

# Data-driven whole mouse brain modeling for multi-scale simulations

Présentée le 17 octobre 2022

Faculté des sciences de la vie  
Institut des neurosciences  
Programme doctoral en neurosciences

pour l'obtention du grade de Docteur ès Sciences

par

**Dimitri Bruno Marie RODARIE**

Acceptée sur proposition du jury

Prof. W. Gerstner, président du jury  
Prof. H. Markram, Dr M.-O. Gewaltig, directeurs de thèse  
Prof. E. D'Angelo, rapporteur  
Prof. T. Leergaard, rapporteur  
Prof. C. Petersen, rapporteur



# Acknowledgements

I could not have achieved this thesis without the help, support, and guidance of many people that I would like to thank here.

First, I would like to thank my thesis director, Pr. Henry Markram for giving me the opportunity to work in the great environment of the Blue Brain Project, and his precious advice to define the subject of my thesis.

I would like to express my many thanks to Dr. Marc-Oliver Gewaltig for his thorough supervision at each step of my work and his guidance to structure my work around scientific questions. His writing advice has been invaluable and will be a great asset for my future career in science.

I cannot thank enough Dr. Csaba Erő for his immense work on the modeling the mouse brain that became the basis of my thesis. He was kind enough to share with me his knowledge and his tools to model the mouse brain from which I could build my own project.

I would also like to thank Dr. Csaba Verasztó, Dr. Yann Roussel, Dr. Daniel Keller, and Dr. Michael Reimann for their precious help on the mouse brain cell atlas project and the fruitful discussions we had on the subject. They helped me to make my assumptions and methods clear for which I am incredibly grateful.

I would like to thank all my colleagues in the Blue Brain Project for creating a friendly and collaborative environment. This environment allowed me to have wonderful scientific discussions and fun moments at the coffee machine. In particular, I would like to thank Ihor Kuras for being a trustable ally, Mike Gevaert for his never-failing engineering support, Jonathan Lurie for our great conversations on the mouse atlas, the whole Machine Learning team for their incredible efficiency and trust, Karin Holm for her invaluable insights in scientific writing and legal procedures. Thank you also Polina Shichkova, Alberto Antonietti, Cyril Favreau, Tristan Carel and Alina Busuioc Jimenez for giving me a reason to come every day to the office with a smile. Finally, I cannot thank enough Stephanie Battini for her infinite kindness and unconditional support, even in the most challenging times.

I would like to thank my close friends and family who supported me during this long endeavor and gave me the strength to achieve this thesis. I would like specially to thank my parents for their unconditional love and for remaining constantly at my side even during the most difficult hours.





# Abstract

In this thesis, we present a data-driven iterative pipeline to generate, simulate and validate point-neuron models of the whole mouse brain. The ultimate goal is to replicate close loop experiments with a virtual body in a virtual world. This pipeline was originally created by a PhD student of the Blue Brain Project and this pioneer work has yielded the first versions of the model and their simulations. My objective is to refine, extend and consolidate the previous version of the workflow and in particular to extend the repertoire of neuron types by integrating and consolidating available literature data.

The pipeline has four main parts. First, we define a pair of reference atlases to create a spatial reference for every dataset that we want to integrate. Second, we create a cell atlas with estimates of the number and density of the major cell types in each brain region. Then, we build a connectivity atlas that describes the connections of each neuron of the mouse brain and their properties. Finally, we assign point-neuron parameters to each of the neuron types, and synaptic parameters to each of the connection types of our model to obtain a point-neuron network of the mouse brain. This network can be simulated to perform mouse brain experiments in-silico.

In this thesis, we evaluate the quality of the different reference atlases, released by the Allen Institute for Brain Science, and provide methods to mitigate the impact of artifacts in the original images. We further extend the cell atlas with density estimates of more inhibitory neuron types. To do so, we introduce a new method to combine estimates of inhibitory neuron counts from the literature into a consistent framework. We also estimate inhibitory neuron density in regions where no literature data are available, using collected literature estimates and gene expression data from in situ hybridization image stacks. Our approach can be further extended to other cell types and provides a resource to build more detailed circuits of the mouse brain. Next, we refine the connectivity atlas, including literature findings on short-range connectivity. Additionally, we construct a database to store data collected on neuron types of point-neuron parameters and synaptic parameters. With the results of the previous steps, we generate an updated version of the whole mouse brain point-neuron network. We tested this new model against its previous version using three benchmark experiments.

## Keywords

Reconstruction, whole-brain, mouse, atlases, cellular composition, connectivity, simulation, point-neurons.

# Résumé

Dans cette thèse, nous présentons un processus itératif, fondé sur l'intégration de données, pour générer, simuler et valider des modèles de réseau de point-neurones du cerveau entier de la souris. L'objectif ultime est de reproduire des expériences en cycle fermé avec un corps virtuel, dans un monde virtuel. Ce processus a été créé à l'origine par un doctorant du Blue Brain Project dont le travail de pionnier a produit les premières versions du modèle et leurs simulations. Mon objectif est d'améliorer, d'étendre et de consolider la version précédente de cette méthode et en particulier d'augmenter le répertoire des types de neurones en intégrant et en consolidant les données disponibles de la littérature.

Le processus comprend quatre parties principales. Premièrement, nous définissons une paire d'atlas de référence afin de créer une référence spatiale pour chaque ensemble de données à intégrer dans notre modèle. Ensuite, nous créons un atlas cellulaire avec des estimations de la densité et les nombres des principaux types de cellules dans chaque région du cerveau. Ensuite, nous construisons un atlas de connectivité qui décrit les connexions de chaque neurone du cerveau de la souris et leurs propriétés. Enfin, nous attribuons des paramètres de point-neurones à chacun des types de neurones et des paramètres synaptiques à chacun des types de connexions de notre modèle pour obtenir un réseau de point-neurones du cerveau de la souris. Ce réseau peut être simulé pour reproduire n'importe quelle expérience du cerveau de la souris *in-silico*.

Dans cette thèse, nous évaluons la qualité des différents atlas de référence publiés par l'Allen Institute for Brain Science et fournissons des outils pour atténuer l'impact des défauts de leurs images d'origine. Nous étendons l'atlas cellulaire avec des estimations de densité d'un plus grand nombre de types de neurones inhibiteurs dans chaque région du cerveau. Pour ce faire, nous adoptons une nouvelle méthode pour combiner les estimations du nombre de neurones inhibiteurs provenant de la littérature dans un cadre cohérent. Nous estimons également la densité des neurones inhibiteurs dans les régions pour lesquelles aucune donnée de la littérature n'est disponible, en utilisant les estimations de la littérature que nous avons rassemblées et les données d'expression génétique provenant d'images d'hybridation *in situ*. Notre approche peut être étendue à tout autre type de cellule et fournit une ressource pour construire des circuits du cerveau de la souris. Par la suite, nous améliorons l'atlas de connectivité, en incluant des résultats de la littérature sur la connectivité à courte distance. De plus, nous construisons une base de données pour stocker les données collectées sur les paramètres point-neurones de chaque type de neurones et les paramètres synaptiques. Avec les résultats des étapes précédentes, nous générons une nouvelle version du réseau de point-neurones du cerveau entier de la souris. Nous comparons ce nouveau modèle à sa version précédente en utilisant trois expériences de référence.

## Mots-clés

Reconstruction, cerveau complet, souris, atlas, composition cellulaire, connectivité, simulation, point-neurones.

# Contents

<b>Acknowledgements .....</b>	<b>3</b>
<b>Abstract .....</b>	<b>5</b>
<b>Keywords.....</b>	<b>5</b>
<b>Résumé .....</b>	<b>6</b>
<b>Mots-clés.....</b>	<b>6</b>
<b>Contents .....</b>	<b>7</b>
<b>List of Figures.....</b>	<b>10</b>
<b>List of Tables.....</b>	<b>11</b>
<b>List of Equations .....</b>	<b>12</b>
<b>Chapter 1    Introduction .....</b>	<b>13</b>
1.1    State of the art for whole-brain modeling.....	14
1.1.1    A workflow to generate simulate and analyze scaffold models of the mouse brain .....	14
1.1.2    Building a cell atlas of the mouse brain.....	15
1.1.3    State of the art on mouse brain connectome reconstruction .....	15
<b>Chapter 2    Reference atlases of the mouse brain.....</b>	<b>17</b>
2.1    Introduction.....	17
2.1.1    AIBS mouse brain reference atlases.....	17
2.1.2    Nissl volume artifacts .....	18
2.2    Methods .....	19
2.2.1    Correct Nissl volume artifacts .....	19
2.2.2    Alignment of annotation volumes.....	21
2.2.3    Alignment of the Nissl volume to the Annotation atlas .....	23
2.2.4    Orientation and Depth .....	25
2.3    Results .....	29
2.3.1    Effect of Nissl refinement on neuron density estimation. ....	29
2.3.2    Annotation volumes realignment .....	30
2.3.3    Orientation and Depth .....	31
2.4    Discussion .....	33
2.4.1    Nissl2 dataset corrections improve cell density estimations .....	33
2.4.2    Realignment of annotation atlases .....	33
2.4.3    Orientation and depth measurements.....	34

<b>Chapter 3</b>	<b>Whole mouse brain cell atlas.....</b>	<b>35</b>
3.1	Introduction.....	35
3.2	Methods .....	36
3.2.1	Overview of the pipeline .....	36
3.2.2	Alignment of in situ hybridization coronal brain slices .....	38
3.2.3	Assumptions underlying the estimation of inhibitory subtypes densities .....	40
3.2.4	Literature review of estimates of the mouse inhibitory neuron densities.....	41
3.2.5	Transfer functions from marker expression to cell density.....	43
3.2.6	Combination of neuron type densities.....	45
3.2.7	Placement of the cells within the mouse brain volume .....	47
3.2.8	Assign me-types to neurons in the whole brain model.....	48
3.3	Results .....	50
3.3.1	An improved computational method to build an atlas of inhibitory neurons in the mouse brain.....	50
3.3.2	Distribution of inhibitory neuron subtypes for the whole mouse brain .....	50
3.3.3	New Excitatory/Inhibitory ratios for the isocortex.....	51
3.3.4	New distribution of me-types for the isocortex .....	53
3.3.5	Validation of the pipeline .....	55
3.4	Discussion .....	58
3.4.1	A new approach to integrate literature findings into the Blue Brain Cell Atlas .....	58
3.4.2	Effect of alignment impact on cell density estimation .....	58
3.4.3	Challenges linked to ISH marker expression datasets .....	59
3.4.4	Limitations of the assumptions of the pipeline .....	59
3.4.5	Reliability of quantitative stereological studies .....	60
3.4.6	Variability of the transfer functions .....	60
3.4.7	Consolidation of neuron density estimates.....	60
3.4.8	Changes in counts of inhibitory neuron in the brain .....	61
3.4.9	Changes in the me-type distribution for inhibitory neurons.....	62
<b>Chapter 4</b>	<b>Whole mouse brain connectivity atlas.....</b>	<b>63</b>
4.1	Introduction.....	63
4.1.1	Long range connectivity reconstruction.....	63
4.1.2	Short range connectivity reconstruction.....	65
4.2	Methods .....	66
4.2.1	Literature based refinement of the local connectivity .....	66
4.2.2	Create a database of Neuron types.....	66
4.3	Results .....	68
4.3.1	Refinement of the cell-to-cell connectivity matrix of the whole mouse brain.....	68
4.3.2	A new database to store point-neuron electrical parameters .....	69

4.4	Discussion .....	71
4.4.1	Impact of the new cell atlas on the whole brain connectivity matrix .....	71
4.4.2	Cell-type recipe-based connectome of the brain .....	71
4.4.3	Storing connectivity rules and electrical parameters into a database .....	71
<b>Chapter 5</b>	<b>Simulation of the whole mouse brain model .....</b>	<b>73</b>
5.1	Introduction .....	73
5.2	Methods .....	73
5.2.1	Point neuron network parameters.....	73
5.2.2	Resting state activity of the SSCtx-LL.....	74
5.2.3	Response to Stimulus in SSCtx-LL.....	75
5.2.4	Whole mouse brain simulation at resting state .....	75
5.3	Results .....	75
5.3.1	Resting state activity of the somatosensory cortex lower limb .....	76
5.3.2	Activity of the somatosensory cortex lower limb stimulated by the thalamus.....	78
5.3.3	Whole mouse brain activity .....	79
5.4	Discussion .....	82
5.4.1	Point-neuron network parameters .....	82
5.4.2	Simulations of the point-neuron mouse brain model .....	82
<b>Chapter 6</b>	<b>Discussion .....</b>	<b>85</b>
6.1	Work achieved .....	85
6.1.1	Analyses and corrections of the reference atlases.....	85
6.1.2	Refinement of the whole mouse brain cell atlas.....	86
6.1.3	Refinement of the local connectivity in the whole mouse brain.....	87
6.1.4	Simulations of the whole mouse brain point neuron model.....	87
6.2	Impact of our pipeline changes on the whole mouse brain model .....	87
6.3	Assumptions of the whole brain model pipeline .....	88
6.4	Challenges related to a data-driven integration of literature data.....	89
<b>Supplementary materials.....</b>	<b>91</b>	
Review of literature on densities of inhibitory neuron in the mouse brain .....	91	
Supplementary figures .....	93	
Supplementary tables .....	102	
<b>References.....</b>	<b>105</b>	

## List of Figures

Figure 1-1 <b>Point neuron whole mouse brain modeling workflow.</b> .....	13
Figure 2-1 <b>Comparison of different versions of the AIBS reference volumes.</b> .....	18
Figure 2-2 <b>Illustration of Nissl volume artifacts.</b> .....	19
Figure 2-3: <b>Corrections the CCFv2 Nissl datasets artifacts.</b> .....	20
Figure 2-4: <b>Alignment of the AV2 to the AV3.</b> .....	22
Figure 2-5 <b>Impact of the quality of alignment between Nissl and annotation volumes on the estimation of cell densities.</b> .....	24
Figure 2-6: <b>Manual realignment of the cerebellar cortex.</b> .....	25
Figure 2-7 <b>Cortical orientation field using Erö's method.</b> .....	26
Figure 2-8 <b>Orientations field and depth computation of the barrel cortex.</b> .....	28
Figure 2-9: <b>Impact of artifacts corrections on the CCFv2 Nissl volume on cortical neuron densities.</b> .....	30
Figure 2-10: <b>Sagittal views of the merged CCFv2 annotation atlas at different steps of the realignment algorithm.</b> .....	31
Figure 2-11: <b>Orientation field of the isocortex and cerebellar cortex.</b> .....	32
Figure 3-1 <b>New workflow for the Blue Brain Mouse Cell Atlas (BBCAv2).</b> .....	38
Figure 3-2 <b>AIBS GAD67 ISH experiment data</b> .....	39
Figure 3-3 <b>Overview of published data on inhibitory neuron densities in the mouse brain.</b> .....	42
Figure 3-4 <b>Linear fitting of marker intensity to cell density in cerebellum, isocortex, and the rest of the brain.</b> .....	44
Figure 3-5 <b>Results of the density pipeline.</b> .....	47
Figure 3-6 <b>Overview of Roussel et al. pipeline.</b> .....	49
Figure 3-7 <b>Distribution of the GABAergic neurons in the mouse brain.</b> .....	50
Figure 3-8 <b>Cortical excitatory/inhibitory ratios.</b> .....	53
Figure 3-9 <b>Distribution of neurons in the BBCAv2 Somatosensory cortex lower limb.</b> .....	54
Figure 3-10 <b>Impact of the amount of literature on the BBCAv2.</b> .....	56
Figure 3-11 <b>Qualitative validation of the BBCAv2 and comparison with BBCAv1.</b> .....	57
Figure 4-1 <b>AAV tracer injection coverage from Oh et al. [29].</b> .....	64
Figure 4-2 <b>Comparison of somatosensory cortex lower limb connectivity matrices.</b> .....	65
Figure 4-3: <b>Database schema for a point-neuron network database.</b> .....	67
Figure 4-4: <b>Generated SSCTX-LL connectivity matrices of the refined whole mouse brain model.</b> .....	68
Figure 4-5: <b>Analyses of the generated cortical connectivity within the whole mouse brain model.</b> .....	69
Figure 4-6: <b>Simulated spontaneous cell-type specific activity.</b> .....	70
Figure 5-1: <b>Resting state spiking activity comparison in SSCTX-LL between the WBM1 and WBM2.</b> .....	76
Figure 5-2 <b>Comparison of the variability of the spiking activity of the SSCTX-LL at resting state.</b> .....	77
Figure 5-3: <b>Comparison of the spiking activity of the SSCTX-LL stimulated by the VPM between the WBM1 and WBM2.</b> .....	78

Figure 5-4: Efferent synapses of the thalamus VPM targeting the SSCtx-LL, sorted by postsynaptic m-type. ....	79
Figure 5-5: Propagation of spiking activity in the whole mouse brain model. ....	80
Figure 5-6: Firing rate comparison for the different brain regions of the whole brain model. ....	81
Sup. Figure 1: Global impact of the annotation atlas on cells and neuron densities. ....	93
Sup. Figure 2: Cortical orientation field computed from each cortical subregion individually. ....	94
Sup. Figure 3: Algorithm 1: Cap inhibitory densities to number of neurons.....	95
Sup. Figure 4: Algorithm 2: Maintain inhibitory densities coherence.....	96
Sup. Figure 5: Cortical inhibitory density distribution.....	97
Sup. Figure 6: Resting state firing rate comparison for the different e-type populations of SSCtx-LL between the WBM1 and WBM2. ....	98
Sup. Figure 7: Spiking activity of the excitatory and inhibitory populations of the whole mouse brain model. ....	99
Sup. Figure 8: Afferent and efferent connections of the SSCtx-LL. ....	99
Sup. Figure 9: Neuron type distribution in the BBCAv1 Somatosensory cortex lower limb. ....	100
Sup. Figure 10: Generated neuron-to-neuron connectome of the refined whole mouse brain model. ....	101

## List of Tables

Table 2-1: Mean intersection over union of the AV2' annotation atlas at different steps of the realignment algorithm. ....	31
Table 3-1 Summary of papers reviewed sorted by marker. ....	41
Table 3-2 Result of the linear fitting. ....	45
Table 3-3 Estimates of inhibitory neurons in the mouse brain in millions and percentage (%) subtypes. ...	51
Table 4-1 Cell atlas neuron counts based on efferent connection types ....	63
Table 4-2 Neuron coverage of the AIBS AAV tracer injections.....	64
Table 4-3: Updated cell atlas neuron counts based on efferent connection types. ....	66
Sup. Table 1: Non standard abbreviations.....	103
Sup. Table 2: Results of the transfer function fitting.....	104

## List of Equations

Equation 2-1.....	21
Equation 2-2.....	22
Equation 3-1.....	40
Equation 3-2.....	40
Equation 3-3.....	43
Equation 3-4.....	45
Equation 3-5.....	45
Equation 3-6.....	45
Equation 3-7.....	45
Equation 3-8.....	45
Equation 3-9.....	46
Equation 3-10.....	46
Equation 3-11.....	46
Equation 4-1.....	63
Equation 5-1.....	73
Equation 5-2.....	74
Equation 5-3.....	74
Equation 5-4.....	74
Equation 5-5.....	75
Equation 5-6.....	75
Equation 5-7.....	75
Equation 5-8.....	75



# Chapter 1 Introduction

The brain is one of the most complex biological structures. Its complexity comes from billions of years of evolution and has been extensively studied over the past century. The Blue Brain Project (BBP) is tackling the tremendous task of reconstructing the brain *in-silico* [1]. The BBP has been creating models of small pieces of rodent brain tissue, capable of reproducing with great details the structure and activity of the brain. To do so, the BBP follows a data-driven approach which consists in integrating experimental data from the literature into coherent frameworks and using this data as parameters of the models, without fitting it, to reproduce experiments. Any discrepancy with the experiment results triggers the refinement of the model and the integration of more literature data. The Blue Brain Project targets to reconstruct the whole mouse brain with great level of details, which includes cell morphology reconstructions, a neuron-to-neuron whole brain connectome and large-scale simulations. This endeavor requires therefore the integration a formidable amount of data from the literature on the mouse brain. The mouse is one of the most studied species and research on its brain has been widely conducted over the past century. Unfortunately, the literature on the mouse brain is incomplete and disparate. There is a great variability in results on the mouse brain due to inter-subject variability (such as its age or sex) but mostly differences in acquisition techniques. More, some results from the literature might be based on old techniques or too specific, for instance targeting a certain region of the brain or a certain genetic breed of the mouse. It is therefore almost impossible to thoroughly evaluate each value from literature and assess their quality. The BBP needs therefore for a strategy to integrate disparate literature data into coherent frameworks to build models of the mouse brain.

The work presented in this thesis is part of the Blue Brain Project and explores methods to construct a data-driven null model of a whole mouse-brain model at the level of point neurons, that is at an intermediate level of complexity that does not incorporate the dendritic and axonal arborizations of neurons. The goal of this project is to provide a scaffold model of the mouse brain in which the detailed regional reconstructions of the BBP could be integrated and have meaningful interactions with the other brain regions. In particular, this work focus was to provide new methods to simplify the integration of literature data and evaluate its quality. Since developing and simulating a model of the whole mouse brain is beyond the scope of a single person, this work draws heavily on contributions from the Blue Brain Project and the Allen Institute for Brain Science (AIBS).

Modeling and simulating the mouse brain is an iterative process. It starts with a coarse initial model that tries to integrate and consolidate enough of the available data for an initial simulation. Then, the model is refined over many iterations: Each iteration starts with a comparison of the model to experimental results. Then, the model gets corrected and refined and more data is integrated and consolidated. Sometimes, the model will reveal gaps in our understanding and will suggest new experiments that need to be done to fill those gaps. To enable an iterative refinement of our model, we want to devise a semi-automatic workflow that generates a model by integrating anatomical and electro-physiological data.

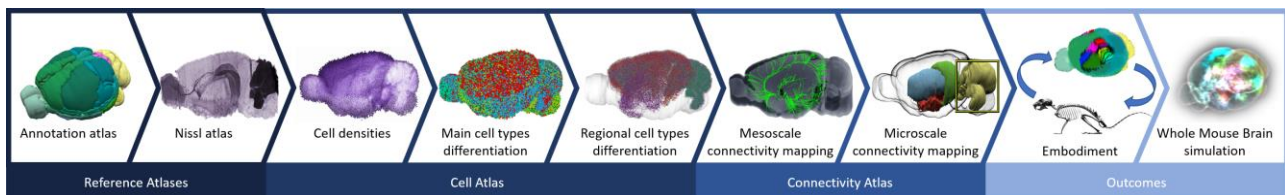


Figure 1-1 **Point neuron whole mouse brain modeling workflow.**

This figure shows the workflow to build a whole mouse brain model. Each stage is represented in different shades of blue at the bottom of the process. Each stage is further divided into steps corresponding to sub projects that will be further described throughout this thesis.

Figure 1-1 shows the workflow that I will be using throughout this thesis was originally designed by Erö [2]. I will present the work that I have done to extend, consolidate and refine the work of Erö. The pipeline has four main steps. The first three steps correspond to the construction of three different atlases. A reference volume that defines the 3D boundaries of each brain region, a cell atlas that prescribes the cell types and their numbers in each brain region, and a connectivity atlas that describes how neurons within and between brain regions are connected to each other as well as the properties of these connections. The last step deals with the

simulation of a whole-brain model, that is with embedding the brain model into a body model and simulating the model using appropriate software tools.

## 1.1 State of the art for whole-brain modeling

In recent years, several authors have published models of entire brains by combining and connecting model neurons in a way that qualitatively resembles a mammalian brain (see review by Stiefel and Brook [3]). However, none of these proposed a reproducible workflow to generate and iteratively refine models, based on large-scale experimental data, such as the approach presented here.

In short, most approaches to modeling the brain (or brain regions) follow one of two approaches. The first approach is called bottom-up modeling. Bottom-up models are constructed from some basic constituents (e.g., the neurons) upwards to understand the macroscopic behavior of circuits constructed in this way. The most prominent example of the bottom-up approach are the circuit models published by the Blue Brain Project [4], [5]. With this bottom-up approach, neuroscientists were able to reconstruct brain tissues with great detail. Unfortunately, it is more difficult with this approach to re-use the knowledge obtained on a piece of the brain in another brain region, as each of them have unique roles and properties.

The second approach is the top-down approach. Top-down models try to implement one or more high-level functions of the brain, ignoring most of the anatomical and physiological structures that give rise to these functions. This approach tries to understand the role of each brain region and how their interactions allow the brain to process information and command the body. The most prominent example of a top-down model is probably SPAUN [6]. However, as the top-down approach focuses on the functions of the brain, it is usually more difficult to link any of these functions to structural or physiological properties of cells or regions.

The work presented in these thesis draws a bridge between the top-down and bottom-up approaches, by creating a point-neuron scaffold of the whole mouse brain. This model is meant to be a reference of the structure of the brain and a placeholder for any property of the mouse brain. It should be used as an anchor point for bottom-up approaches to place their models in the greater context of the whole brain. On the other side, it offers a reference of the structure of the network within the brain that can help top-down approaches to attach the functions of the brain to biological properties. Our model and the pipeline behind it should therefore be able to integrate heterogeneous datasets from literature in a coherent framework. Moreover, it should provide a default solution in absence of data. Finally, the model should offer tools to analyze the structure, properties, and activity of the brain network in order to define necessary corrections to implement in the model. In this cycle of integration of new data, refinement of the model and the methods, simulation, and analysis of the results, we ensure that our model will improve with each iteration. Our approach is data-driven by definition, which means that we are not fitting our parameters to obtain specific features but rather we let the features emerge from the constraints defined by the data. This work fits into the Blue Brain Project (BBP) roadmap as it provides a reference placeholder that will be refined with the detailed reconstructions of brain tissues that have been done in the laboratory. More precisely, the pipeline fits in the BBP most recent milestone that targets the creation of a whole mouse brain model.

### 1.1.1 A workflow to generate simulate and analyze scaffold models of the mouse brain

The goal of this work is to provide a data-driven iterative process to generate, simulate and validate point-neuron whole brain models of the mouse with the ultimate goal to replicate close loop experiments with a virtual body in a virtual world. Our workflow comprises several stages corresponding to the different atlas models, shown in Figure 1-1.

The process starts with the definition of a pair of Reference Atlases: an annotation atlas and a Nissl volume. These two brain volumes are used as spatial reference to realign each dataset of the pipeline. The annotation atlas associates a brain region to each voxel of the brain, which allows for instance, to estimate the volume of brain regions. The Nissl atlas is a three-dimensional brain volume, reconstructed from an aligned image stack of Nissl-stained brain slices [7]. The AIBS Nissl volume is derived from the original experiment from Dong [7], which means that it suffers from the artifacts of the original dataset. Original images have been torn and contain artifacts in the Nissl intensity due to holes or over/under exposure. These artifacts will have an impact on the cell density estimations. Hence, I worked on several tools to improve the quality of these datasets. These tools will be described in the following chapter.

Nissl stains primarily cell somas. This property was used by Erö et al. [8] to estimate the cell densities in the whole mouse brain and build the first whole mouse brain cell atlas (BBCAv1). This cell atlas describes the cellular composition in the whole mouse brain. Its construction proceeds in a coarse-to-fine approach. Each step refines the model by providing a more fine-grained cell type to the model. In the end, the cell Atlas provides a position and a cell type label to each cell in the model. In the third chapter, I will describe the work to refine the Cell atlas of the mouse brain. In particular, this chapter will focus on a new method to integrate heterogeneous

and sometimes contradicting data from literature in a consistent framework. We will produce a new version of the Cell Atlas which integrates more cellular types and facilitates its future extension.

Then, we build a connectome of the mouse brain registering the connections between each neuron of the Cell atlas: a connectivity atlas. This model uses Adeno-associated virus (AAV) tracer injections datasets from the AIBS for long-range connectivity and kernel-based connectivity for short-range connectivity. In Chapter 4, I will propose a new solution to integrate connection type properties. Literature provides indeed some information about how the diverse types of neurons in the brain connect to each other. We need a method to store this information and retrieve it while creating the connectome of the brain.

The combination of the Cell atlas and the Connectivity atlas allows the creation of a point neuron network of the whole mouse brain. This is done by assigning a set of electrical parameters to each neuron of the Cell Atlas and similarly synaptic parameters to each connection of the Connectivity Atlas. Once a new version of the whole brain model is created, we should be able to benchmark it in simulations and compare its results to the previous instance of the model created. In Chapter 5, I will present three benchmark experiments that I used to test the new whole brain model.

### 1.1.2 Building a cell atlas of the mouse brain

Over the past century, numerous studies have reported a vast variety of cells in the mouse brain according to their morphological, electrical, and molecular features. Several groups such as the BRAIN Initiative® have launched ambitious projects to undertake a comprehensive census of all brain cells, determining their molecular, structural, and functional properties [9]–[12]. Building a catalog of the vast diversity of neuron types in the brain is a challenge for modern neuroscience. Characterizing this variety is exacerbated by the fact that distinct cell types are localized to specific brain regions - for example, pyramidal cells in the cerebral cortex or Purkinje cells in the cerebellum [13], [14]. Therefore, studies tend to simplify this task by focusing on specific brain regions only. Moreover, the variety of methods used results in considerable variability in the outcomes, even when the same region is considered [15]–[17]. The cellular composition problem can only be conclusively solved by integrating disparate datasets from literature into a coherent framework. Obtaining a complete and comprehensive cell atlas of the mouse brain is, therefore, a monumental task, which needs to be tackled to enable the *in-silico* reconstruction of multi-scale neural circuits [4], [14], [18]–[20].

Erö and colleagues [8] were able to estimate the cellular composition of each brain region defined by the Allen Mouse Brain Atlas [7] with respect to excitatory (glutamatergic) and inhibitory (GABAergic) neurons, astrocytes, oligodendrocytes, and microglia. The authors aligned multiple datasets, including genetic marker expression from *in situ* hybridization (ISH) experiments from the Allen Institute for Brain Science (AIBS) to the Allen Reference Nissl volume from Lein et al. [21], to determine the positions of all cells within the annotation volume. The authors used global constraints from the literature [17], such as the total number of cells and neurons in the mouse brain from Herculano-Houzel et al. [22] to avoid any bias toward a particular region. This resulted in the first version of the Blue Brain Mouse Cell Atlas pipeline (BBCAv1) as described in Erö et al. [8].

### 1.1.3 State of the art on mouse brain connectome reconstruction

Once the cellular composition of the whole brain is defined, we need to describe how these cells connect to each other to form the brain network. In this thesis, we will focus on the neural circuitry as the impact of glia cells on neuronal activity is not well defined as for now. Neurons connect to each other mostly by creating synapses on the apposition of the axon of the source neurons on the dendrites of the target neurons. Multiple studies have described the whole brain connectome, whether defining it with functional or structural connectivity and at different scales. Reconstructing the brain connectome is a complex task due to the wide variety of cell and connection types that compose it. Some approaches attempt to reconstruct the connectome of the brain from a functional point of view [23]–[26]. These approaches leverage *in-vivo* brain imaging techniques to deduce connections between regions of the brain based on its activity during resting-state or specific task experiments. For instance, Allen et al. [23] uses functional magnetic resonance imaging to study the macro-scale spatio-temporal activity of the brain. From these images, the authors could analyze time-correlations between regions of the brain which are translated into connections between these regions. Local motifs of correlated activity have also been studied using more precise imaging techniques such as Calcium [25] or Voltage Sensitive Dye imaging [26]. These techniques provide a lot of information on the macro-scale to the meso-scale structure of the brain. However, these *in-vivo* imaging techniques highlight correlation of activity which sometimes does not translate to actual direct synaptic connections. A correlation of activity between a pair of regions might be the result of the activity in a third region, which might be difficult to prove without knowing the underlying structure and the delay of propagation between regions of the brain. Moreover, the spatial resolution of the signal images makes it difficult to deduce any cell-to-cell connectivity. The wide range of electrical parameters involved during brain activity cannot be deduced solely based on brain activity without overfitting.

Other studies leverage morphology reconstructions of the neurons to find appositions between their arborizations once placed in the brain volume [4], [27]. Morphology appositions are pruned to extract synapses which gives the complete connectome of the circuit reconstructed. However, neuron morphologies are not known in every region of the brain and their axon projections are usually not reconstructed as it can spread across multiple regions of the brain. Several groups work on the reconstruction of long morphologies that spread across the whole mouse brain [28]. Unfortunately, the hundreds of high-quality reconstructions that this endeavor could produce are sparsely covering the whole brain and will not be sufficient to reconstruct a whole brain connectome.

Other studies use markers to highlight the arborization of neurons of the brain. Among them, recombinant adeno-associated viruses (AAV) are widely used to infect a population of neurons in the mouse brain and make them express green fluorescent protein (GFP) in their dendritic or axonal projections. Oh et al. [29] performed thousands of anterograde rAAV injections in order to highlight axonal projections in the mouse brain and reconstruct its mesoscale connectivity matrix. The rAAV virus infects every neuron in its injection vicinity which means it is difficult to link the axonal projections highlighted to a specific neuron population. Some studies therefore use modified rAAV tracers that target specific neural populations [30]. The anterograde rAAV tracer highlights the entire axonal arborization, and when injected in a population of neurons, it is sometimes difficult to predict in which parts of the brain these axons create actual connections. A solution for this issue is to inject both an anterograde and a retrograde virus in the predicted target region of the projection studied [31]. The mesoscale connectivity matrix obtained through these studies is however difficult to translate into cell-to-cell connectivity matrix as the tracers injected infect multiple neurons and do not give any information on the electrical properties of the connections it highlights.

The electrical properties of these synapses can be studied using multiple patch clamp recordings and modeled with great details. Using this technique, the complete connectome of micro-organisms could be reconstructed [32]. However, each neuron of the mouse brain can create up to thousands of synapses, each synapse with its own set of parameters. Reconstructing the whole brain connectome by analyzing each neuron individually would be a task too long and tedious to be achieved in decent times. Studies tend to regroup synapses into types according to the source and target neurons they connect [4], [14].

# Chapter 2 Reference atlases of the mouse brain

## 2.1 Introduction

The whole mouse brain pipeline is defined according to a pair of reference volumes: an annotation atlas and a Nissl volume. These volumes are used to align other mouse brain datasets from the AIBS to a common vector space. Examples are *in situ* hybridization experiments to determine gene expression patterns or the rAAV tracer studies to determine the mouse meso-scale connectome [21], [29]. Since the vector space is common for all datasets, it has been termed: the Common Coordinate Framework (CCF). Additionally, the reference atlases can be used to compute cell orientations and depth within brain regions, which are required information to place neuron morphologies in the mouse brain. In this thesis, we will use the reference datasets from the AIBS as the basis for our cell atlas and whole-brain model, as they are widely used by the neuroscience community.

Over the years, the Allen Institute has produced multiple versions of the Common Coordinate Framework, each with its own set of advantages and disadvantages. The first version of the CCF still had a jagged and noisy reference volume but was closely aligned to a stack of Nissl-stained images from a single individual. Later version of the CCF had a much smoother reference volume, which was derived by averaging data from many individuals. However, as a result, the correspondence to the Nissl data of one individual was lost. In this chapter, we will explore different approaches to adapt these reference atlases for our purposes and provide methods to choose the best pair of reference atlas to estimate cell densities in all annotated brain regions.

### 2.1.1 AIBS mouse brain reference atlases

Each AIBS Nissl reference volume is derived from high resolution image stacks of Nissl-stained coronal brain slices [7]. These images are realigned and arranged to form a 3D brain volume. Nissl stains each cell somas. The intensity of each voxel of the Nissl volume correlates with the number of cells in this voxel. This Nissl volume has a corresponding annotation volume (AV) which maps each voxel of the Nissl volume to a brain region. The AIBS organize brain regions in a hierarchical tree, with leaves representing the finest parcellation.

The original AIBS Nissl datasets were produced by Dong [7]. The data consists of 25  $\mu\text{m}$  thick coronal slices (Nissl stained) from an adult wild-type mouse brain (P56). High-resolution images of these slices were arranged to form a 3D brain volume. In addition, experts manually traced the contours of each brain region on the right side of each image. The annotations were then mirrored to the left side of the images. To match this new annotation volume the left side of the Nissl volume was also obtained by mirroring the other side. The final images were then down sampled to form the first version of the Nissl volume (Nissl1) and Annotation volume (AV1), published in 2011. Since its first publication, the AIBS has released multiple versions of the Allen Brain Nissl volume and their associated annotation volumes.

At the level of individual slices, Nissl1 and AV1 were well aligned to each other, while adjacent slices were not. This rather led to a jagged Nissl brain volume and annotation volume. Therefore, Erö et al. [8] realigned adjacent Nissl and AV slices along the sagittal axis with a non-rigid algorithm to create a new reference pair  $\text{Nissl}_{\text{bbp}} + \text{AV}_{\text{bbp}}$ , the CCFbbp (see Figure 2-1A). This annotation volume ( $\text{AV}_{\text{bbp}}$ ) labels 955 unique region id including fibers tracts and is the basis for the Blue Brain Cell Atlas. The resulting brain volume is still jagged despite realignment and contains some holes due to the absence of the annotations for the ventricles (these were added in 2015). Additionally, some voxels of the  $\text{AV}_{\text{bbp}}$  are not assigned to their closest leaf region.

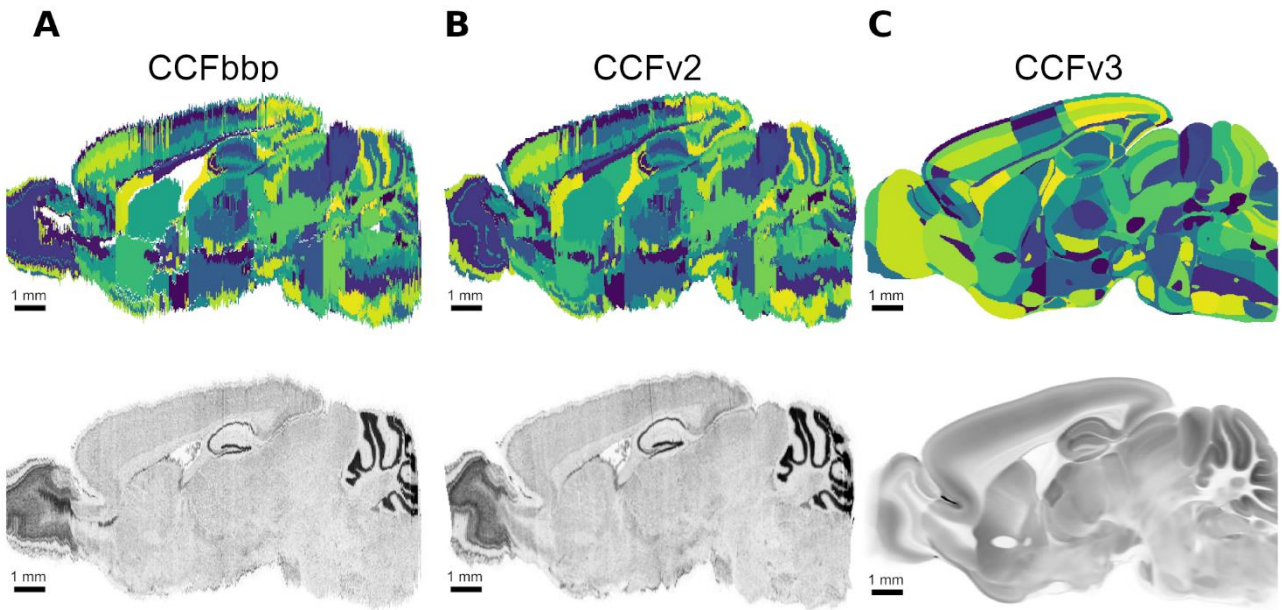


Figure 2-1 Comparison of different versions of the AIBS reference volumes.

Different versions of the reference volumes are shown in sagittal view. Top row: annotation volumes and bottom row: corresponding Nissl volumes. Brain regions in the annotation volumes are shown in arbitrary shades of green. (A) The CCF<sub>bbp</sub> annotation and Nissl<sub>bbp</sub> volume. (B) The CCFv2 annotation and Nissl2 volume. (C) CCFv3 annotation volume and AIBS average brain volume. While all versions are based on the original Nissl1 + AV1 AIBS datasets released in 2011, significant differences can be observed: CCFv3 (C) has the smoothest annotation and brain volumes. However, the version has no associated Nissl volume as in (A) and (B) and the average brain volume from which it has been derived cannot be used for cell density estimation.

The AIBS used a comparable technique to obtain the CCFv2 (see Figure 2-1B), comprising Nissl2 + AV2. Both Nissl1 and Nissl2 are obtained from the same individual, but Nissl2 is aligned to an averaged mouse brain volume (see Figure 2-1C bottom) that was obtained from over a thousand adeno-associated virus (AAV) tracer experiments [29], [33]. AV2 has 963 unique region ids, including fibers and ventricles. Both AV<sub>bbp</sub> and AV2 still have jagged edges and some voxels do not belong to a leaf region in the region hierarchy. The AIBS then published a third AV version (AV3) that was completely derived from this average brain (see Figure 2-1C), using the pipeline described in Wang et al. [34]. While AV3 is the smoothest available AV, it no longer has an associated Nissl volume; it is paired with the closest Nissl volume available (Nissl2). AV3 has only 839 unique region ids, including new region ids for the isocortex, thalamus and brainstem. However, some important structures such as the molecular and granular layers of the cerebellum are no longer identified in AV3 but have been merged.

### 2.1.2 Nissl volume artifacts

Since the Nissl volumes were generated from high-resolution images of physical brain slices, they have also captured artifacts from the original microscopy images from Dong [7]. Figure 2-2 illustrates some of these artifacts, including differences in exposure levels during image acquisition (arrows in A) as well as bubbles, tissue tears (arrows in B). To quantify exposure differences between adjacent slices, we computed the median Nissl expression along the rostral-caudal axis. Figure 2-2A shows the sagittal projection of the median Nissl expression level. The corresponding medial values are shown for the entire brain in C and for isocortex only in D. The

colored bars indicate significant changes in exposure levels. Sometimes individual slices are exposed so differently from the neighboring slices, that they show up as spikes in the median plots (see arrow in A and spikes at the same section position in C and D).

Figure

2-2

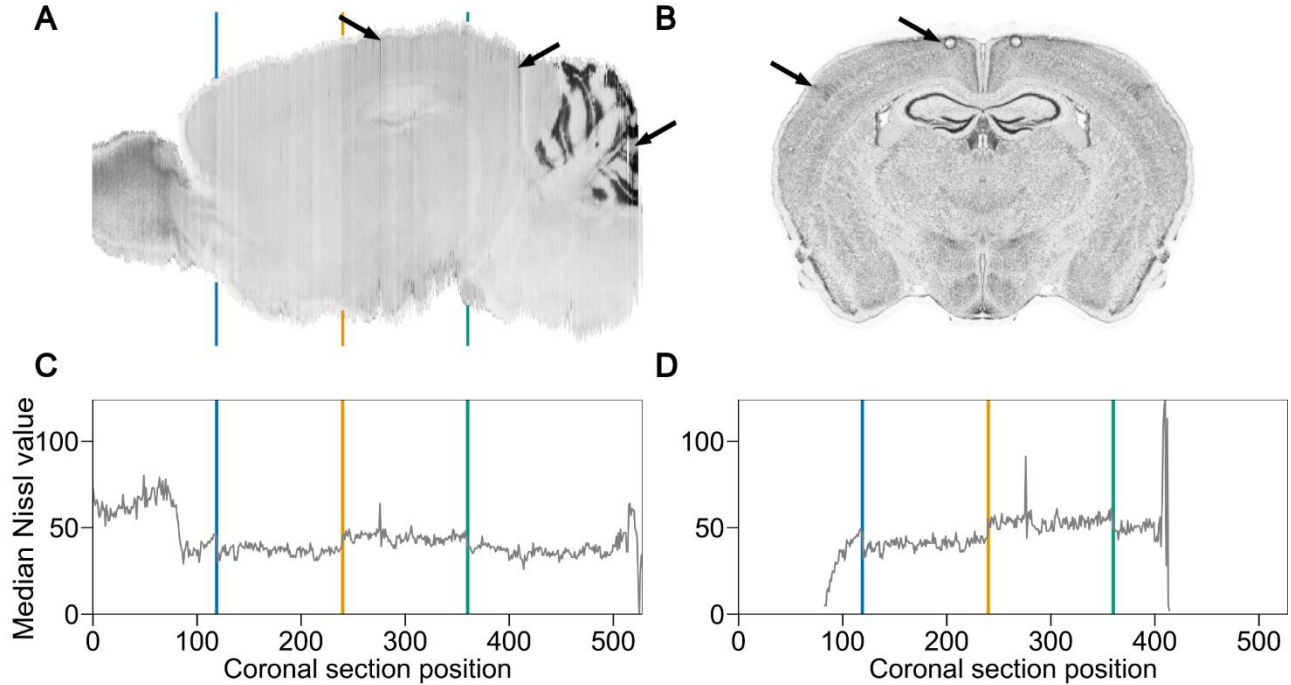


Figure 2-2 Illustration of Nissl volume artifacts.

(A) Sagittal projection of the median expression of the Nissl volume realigned in Erö et al. [8] (Nissl<sub>bbp</sub>). Regions with high Nissl intensity appear in dark gray. The blue, orange and green lines behind the sagittal slice highlight the rapid changes of Nissl expression coming from the original Nissl experiment from Dong [7]. Arrows show single coronal slices with high Nissl expression value in comparison with their neighboring slices. (B) Coronal slice of the Nissl<sub>bbp</sub> reference atlas, showing cells of the mouse brain. The arrows show artifacts in the volume due to image acquisitions: bubbles, shrinking, etc. (C) (D) Evolution of the median expression level along the sagittal axis, in the Nissl, for the whole brain (C) and the isocortex (D). Colored lines indicate the location of the rapid changes of Nissl expression that were detected in (A). These changes translate into an overall rise or drop in expression levels in (C) and (D).

Although annotations were drawn on top of the Nissl images (for CCFbbp and CCFv2), the resulting volumes are sometimes locally not perfectly aligned to each other. For CCFv3, the misalignment is even worse, since AV3 was not derived from a Nissl volume. In particular, AV3 is not aligned to Nissl2. This can lead to errors in the computation of the volumes and cell densities in the different brain regions.

## 2.2 Methods

### 2.2.1 Correct Nissl volume artifacts

In this section, we propose some ways to correct artifacts in the Nissl volume and to measure their effectiveness to improve the estimation of cell and neuron densities. To correct Nissl artifacts, we need some a priori knowledge about the intensity of Nissl in the affected brain regions. We will therefore focus on the isocortex for which we assume that the Nissl intensity of each layer should be relatively constant across regions. We will further assume that the intensity of Nissl in layers 2 to 6 is similar. Our assumptions imply that the Nissl intensity across coronal slices should follow the same distribution. We can therefore use the Nissl distribution in one coronal slice of the isocortex as a reference and match the distribution of all other slices. This corresponds to a histogram-matching algorithm. Here, it will be used to compensate for the exposure differences detected in Figure 2-2. As reference, we chose the coronal slice with the most voxels labeled as a cortical region (see Figure 2-3A). Since this slice is within the range of reduced expression levels (see Figure 2-2ACD between blue and orange markers), histogram matching will reduce the level of Nissl expression for isocortex in



slices outside this range. However, the total estimated numbers of cells and neurons in the isocortex should not be affected, since Erö et al. algorithm [8] tries to match specific targets from Herculano-Houzel et al. [22]. Figure 2-3A shows the selected reference slice. It not only has the largest proportion of cortical voxels, but it is also a slice with very few artifacts, which makes it a good reference for histogram matching. The annotations of L1 and L2-L6 are indicated in blue and orange, respectively. We computed the histogram of Nissl expression in L1 and L2-L6 in the reference slice and matched the distribution of the corresponding layers in all other slices. Figure 2-3B shows the Nissl intensity distribution for L1 (top) and L2-L6 (bottom). We notice that the distribution of the Nissl intensity in L1 has a long right tail, indicating voxels with high Nissl intensity values. These voxels are likely artifacts and should not appear in a layer with low cell density. In fact, as we can see in Figure 2-3A, some voxels of L2 with higher cell density (darker because denser in cells) have been mislabeled as L1. As we will discuss in Section 2.2.3, these errors have a significant effect on cell density estimation. Unfortunately, histogram matching will propagate these artifacts to other coronal slices. We applied the histogram matching algorithm to the isocortex (see Figure 2-3C) and obtained a new Nissl volume (see Figure 2-3D middle panel)

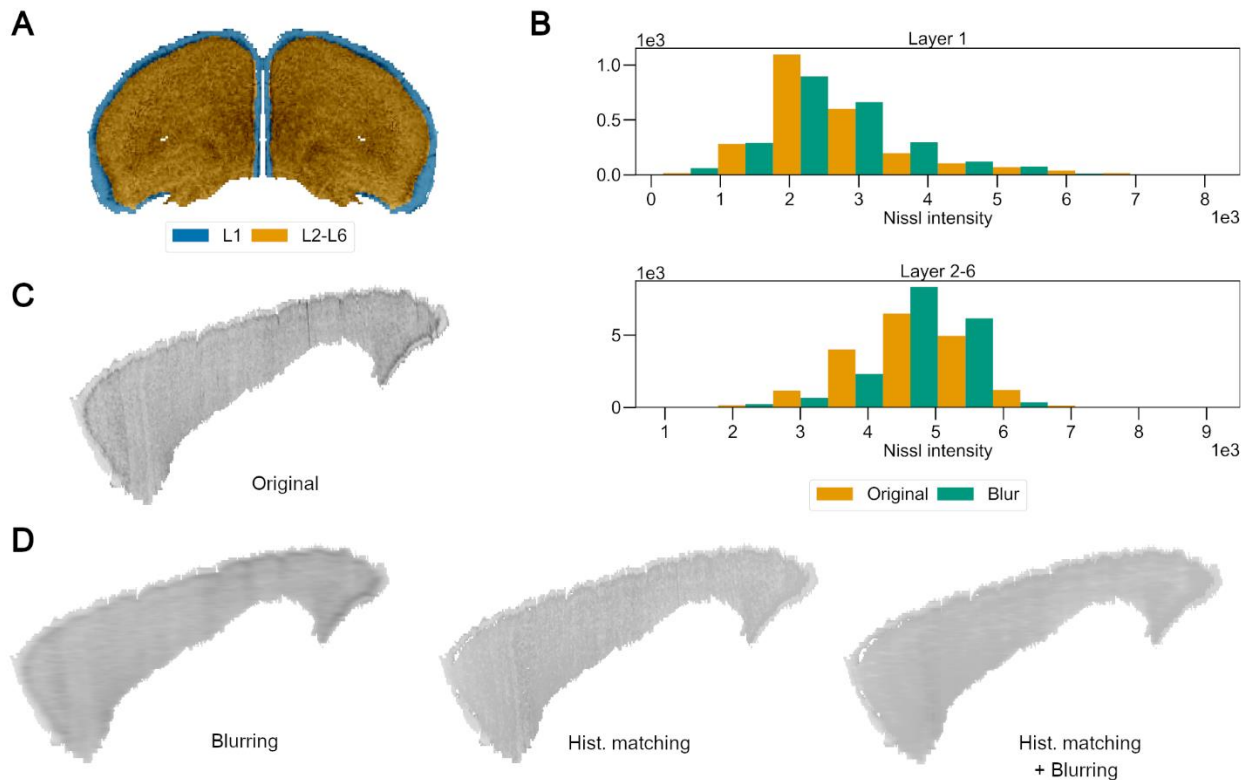


Figure 2-3: **Corrections the CCFv2 Nissl datasets artifacts.**

(A) Coronal cortical Nissl slice used as reference for histogram matching (slice position=138). The CCFv2 Nissl expression is shown in levels of gray and the annotations are overlaid in colors.

(B) Histograms of the Nissl intensity in the coronal cortical reference slice (A) in the original and blurred CCFv2 Nissl volume. These histograms were normalized to estimate the empirical distribution of the Nissl intensity in the reference slice. Then, this distribution was applied to all other coronal slices.

(C) Sagittal section of the original CCFv2 Nissl volume showing the isocortex only.

(D) Sagittal section of the CCFv2 Nissl volumes obtained after blurring (left panel), histogram matching (middle panel) and combination of the two methods (right panel).

To remove local perturbations of the Nissl expression, we used a Gaussian filter to smoothen the intra-layer Nissl expression. The Gaussian filter was applied to L1 and L2-L6 separately and isolated from the rest of the brain. We chose a low standard deviation for the filter (here 3 voxels) to maintain potential inter-region differences. The Gaussian filter was applied along the sagittal direction as local artifacts are usually located in a single coronal slice (see results in Figure 2-3D left panel).

We also tested the combination of Gaussian filtering (blurring) and histogram matching (see results in Figure 2-3D right panel). The effects of these methods on the Nissl dataset can be compared in Figure 2-3CD.



## 2.2.2 Alignment of annotation volumes

The AV3 is the smoothest AV available and should therefore be the best candidate to build models of the mouse brain. However, AV3 lacks some important subregions that were still present in AV2 (e.g., molecular and granular layers of subregions of cerebellar cortex). In order to obtain the most detailed and smoothest AV, we want to realign the AV2 to AV3 so that we can combine them to obtain the most complete region description. In this chapter, we will present an approach to improve the alignment between the two AV dataset versions.

The algorithm presented here is based on the Advanced Normalization Tools (ANTs) software package [35] as described by Krepl et al. [36]. ANTs allows the registration of an input dataset (image or volume) according to a reference dataset. This is done by computing a displacement field which can be applied to the input dataset to match its reference. Here we want voxels of the AV2 (containing regions id) to match those of AV3.

Since AV2 and AV3 do not use the same set of region ids, the task of realigning their subregions is more difficult. we therefore first re-labelled voxels of AV2 and AV3 so that they can be compared to each other. Voxels labeled with not common ids were set to their closest common parent id. For instance, hippocampus CA1 subregions are not present in AV3 but are in AV2. Voxels in AV2 belonging to these subregions were relabeled to hippocampus CA1. Additionally, some isolated voxels in AV2 were not labeled with the most precise region id. We assigned to these voxels the id of the closest leaf region, using a Breadth-First-Search algorithm. Finally, for the isocortex, we also kept the layer subdivision of the AV when it was possible, as it gives more landmarks for ANTs to work with. For instance, in the visual areas, the AV3 have introduced new regions without any link to the region segmentation in AV2. All L1 regions of the visual areas were hence labeled with the same id and similarly for layer 2 to 6.

This procedure yielded two new AVs (AV2' and AV3') from respectively the AV2 and AV3 (see Figure 2-4AB). These datasets describe the same brain volume with the same set of region ids which means that the AV2' should correspond to a slightly deformed AV3'. These deformations can be captured by ANTs. we used ANTs Symmetric Normalization method which computes linear deformations (translations, rotations) to the entire volume to maximize a similarity between the reference and the moving datasets. However, there are several difficulties that we needed to address:

1. ANTs realignment is based on features detection in the voxels' intensity of the input datasets. However, the AVs' voxel contains the regions' id, and these ids are arbitrarily distributed in the brain. ANTs will give more weight to the boundary between regions with distant ids as it will detect it as a feature. Conversely, the boundary between regions with consecutive ids might not be detected as a feature.
2. The voxels within a region all have the same weight which allows ANTs to move them arbitrarily.
3. After registration, ANTs provides a displacement field of the whole dataset (see Figure 2-4F for an example displacement field), which means that the corrections applied to a small region of the brain affect all other regions.

To simplify the problem, we applied the realignment to the borders of each brain region, because if the borders of the AV2' match the AV3' counterparts, the voxels within the region should do so too. The borders of a region are defined as the set of voxels surrounding the region of interest (see Figure 2-4CD). If the ANTs registration tool were to be applied on the whole brain dataset with all the regions borders, it would be more difficult for the program to determine which border should be realign to which. We therefore only apply the registration tool on isolated borders (i.e., region by region). Finally, we limited the effect of the displacement field obtained after registering a region on the others, by authorizing the displacement on the voxels that belong to the parent region (see example authorized mask in Figure 2-4E). For instance, when realigning SSCTX-LL layer 1, the resulting displacement field should only affect voxels within SSCTX-LL. This means also that SSCTX-LL must be realigned before realigning its layers.

The algorithm therefore follows the region hierarchy from top to bottom. For each region, the algorithm takes the borders of its direct children in AV2' and registers them to their counterparts in AV3'. The displacement field is applied to the region voxels of AV2' while the remaining voxels remain fixed. We repeated this step until the registration converged, which means that the displacement field was not improving the alignment of the AV2' with respect to AV3'. The realignment score for a region  $r$  was defined as the ratio of incorrect border voxels in AV2' according to AV3' after registration:

$$score(r) = 1 - \frac{count(border\_voxel\_AV2'(r) \neq border\_voxel\_AV3'(r))}{count(border\_voxel\_AV3'(r))} \quad \text{Equation 2-1}$$

To measure the quality of the realignment, we also used the Intersection over Union metric (IoU). This metric is commonly used to evaluate algorithms that detect objects in images. The predicted area ( $A_{predicted}$ ) where the object is detected can be directly compared

to the ground truth area ( $A_{reference}$ ). The IoU is defined by the ratio of the area common to the prediction and the ground truth divided by the area of the union of the two areas:

$$IoU = \frac{A_{predicted} \cap A_{reference}}{A_{predicted} \cup A_{reference}} \quad \text{Equation 2-2}$$

We calculated the mean of the IoU of each coronal slice of the AV', comparing the areas occupied by each brain region of the AV3' and the AV2' at each step of the algorithm.

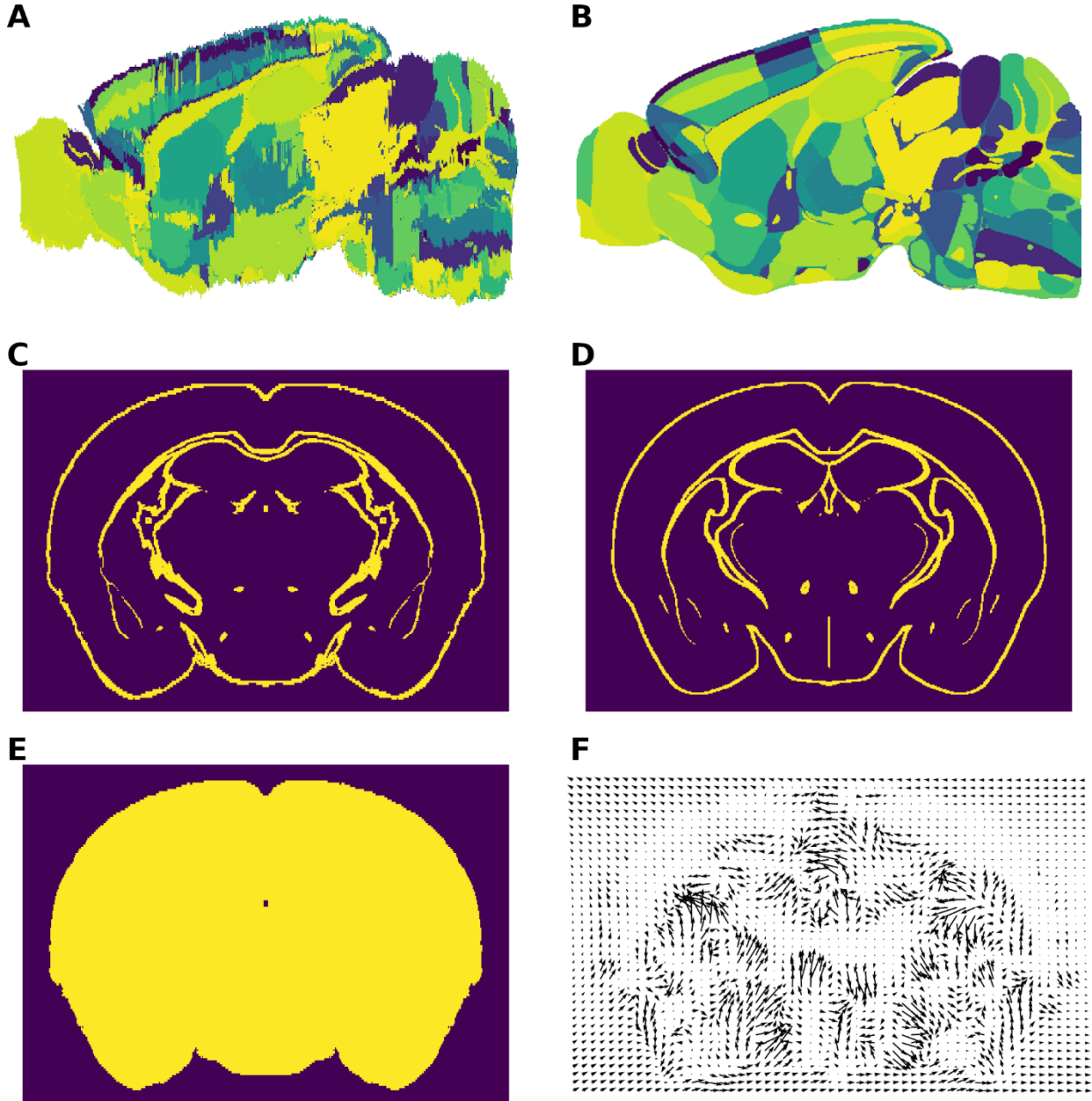


Figure 2-4: **Alignment of the AV2 to the AV3.**

(A) (B) Merged version of respectively the CCFv2 (AV2') and CCFv3 annotation atlas (AV3'). The regions ids of the AV2 were changed to match the ones present in the AV3 and vice versa for the AV3'. These versions of the annotation atlas have the same set of region ids. To each brain subregion of the annotation volume is assigned a random color. (C) (D) Coronal slice of the brain showing the border of the "Basic cell groups and regions" region in respectively the AV2' and AV2'. (E) Mask of the "root" region in the AV2', i.e., the parent region of the "Basic cell groups and regions" region according to the AIBS region hierarchy. (F) Displacement field computed with the ANTs realignment tool to realign the image (B) to the reference image (D). This displacement field, limited to the mask (E), was applied to the AV2' to match the AV3'.

### 2.2.3 Alignment of the Nissl volume to the Annotation atlas

When we estimate cell densities from a Nissl volume using Erö et al.'s method [8], different combinations of Nissl volumes and annotation atlases yield different results (see Figure 2-5CD). In this section, we present our analysis to identify the best reference dataset pair (referred to as CCF version) for density estimation. We find that the quality of cell density estimation strongly depends on how well the Nissl volume is aligned to the annotation volume.

Poor alignment means that parts of one region in the Nissl volume are mislabeled as being part of a neighboring region. This means also, that parts of the cell estimates are assigned to wrong regions. To assess the quality of the alignment we look at parts of the brain where the boundary between two regions is clearly visible in the Nissl volume, specifically the border between layer 1 and layer 2 in the isocortex and the border between the molecular and granular layers of the cerebellar cortex. Figure 2-5AB illustrates the quality of alignment between Nissl and AV volumes for the different reference atlases for these regions.

We apply the Erö et al. pipeline [8], for the three pairs of annotation/Nissl volumes:

1. the CCFbbp version derived from the Nissl1 and AV1, which we name Nissl1a and AV1a.
2. the CCFv2 version (Nissl2 + AV2) from the AIBS also derived from the Nissl1 and AV1. These datasets are publicly available on the AIBS website [37], [38].
3. the CCFv3 version (Nissl2 + AV3), 2017 release version (most recent) also created by the AIBS and accessible on their website [39].

For each pair of reference volumes, we first estimate the densities of cells and neurons per voxel in the corresponding brain volume, based on the total number of neurons given by Herculano-Houzel et al. [22]. We then compare the resulting distribution of cell densities to each other and to the literature (see Figure 2-5CD) [40]–[42]. Despite an overall agreement (see Sup. Figure 1), we observe significant differences for smaller regions. For regions with low cell density such as isocortex layer 1, this difference can scale up to two times more cells for CCFv3 as compared to CCFv2 (see Figure 2-5C), and even more compared to some literature values.

To quantify the impact of such misalignments more precisely, we manually realign AV2 to Nissl2 for the lingula, and flocculus regions of the cerebellum. The cerebellar cortex is composed of three layers with very different cell distribution: the granular layer, the Purkinje layer, and the molecular layer. The Purkinje layer is a single cell thick layer of Purkinje cells. The granular layer is a very dense region containing mostly granular cells which have very small somas. Finally, the molecular layer has a low cell density and contains two populations of inhibitory neurons (Stellate and Basket cells). First, we relabel the voxels of the Flocculus, Crus 1 and Lingula as either molecular layer or granular (+ Purkinje) layer based on their Nissl intensity. Indeed, the significant difference in cell densities between the granular and molecular layer is also present in the distribution of Nissl intensity (see Figure 2-6A). We chose a threshold of Nissl intensity based on user-selected voxels at the observed border between the two regions. Results are shown in Figure 2-6B. The borders of the regions were sometimes incorrect, and some artifacts were still present (see Figure 2-6C). We manually corrected these issues and obtained a new version of the annotation atlas for these regions. The manual correction of the annotations is however a very tedious and time-consuming task, even at a voxel resolution of 25µm. We therefore applied correction solely on the flocculus and lingula.

We find that these corrections indeed lead to density estimates closer to their literature values (see Figure 2-5D).

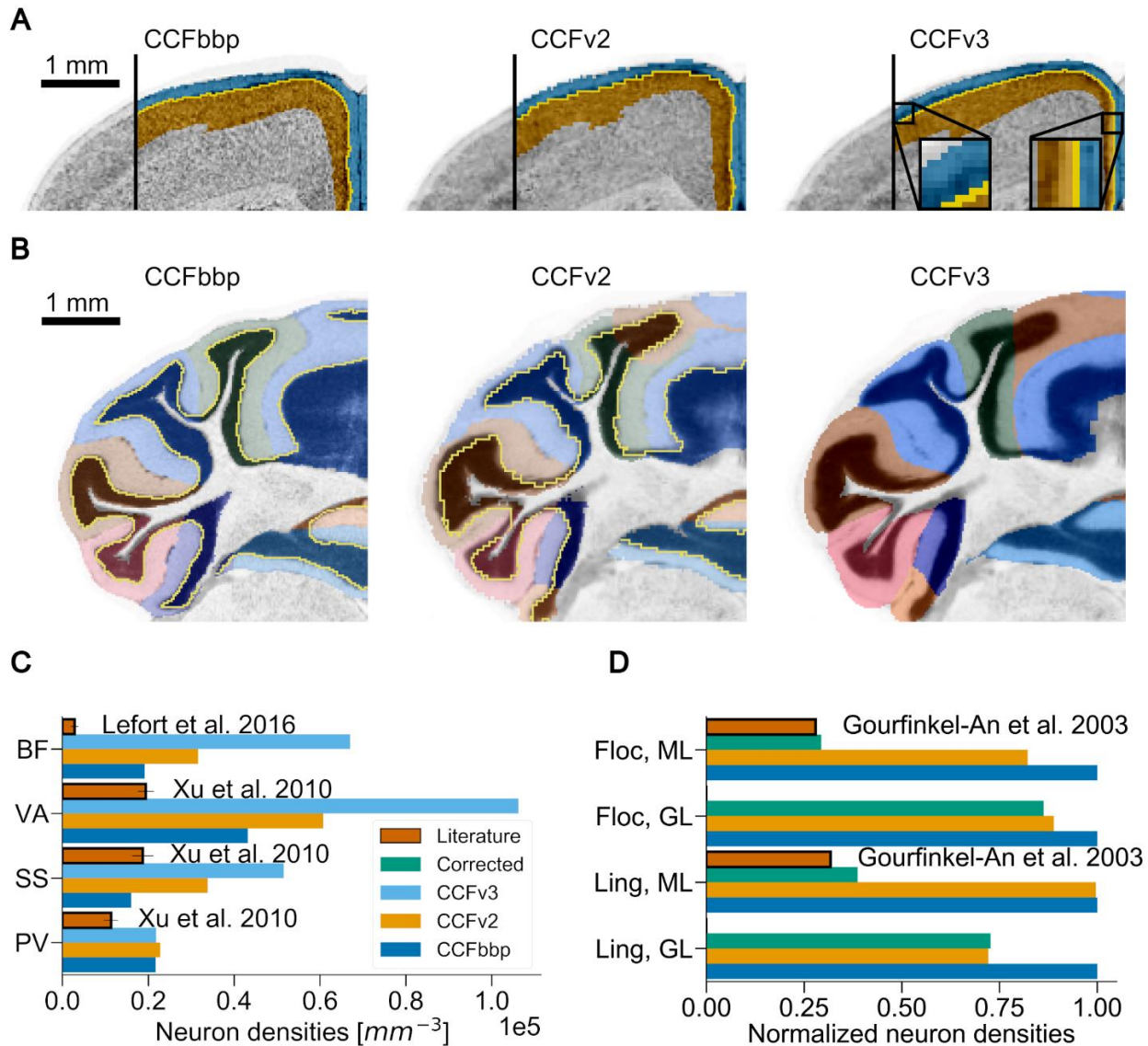


Figure 2-5 **Impact of the quality of alignment between Nissl and annotation volumes on the estimation of cell densities.**

Panels (A) and (B) show the precision of alignment between the Nissl datasets in gray and their respective annotation atlases overlaid as colors (coronal slices). Boundaries between annotated regions, when available, are shown in yellow. The different CCF versions are shown from left to right: CCFbbp, CCFv2 from the AIBS, and CCFv3 from the AIBS. Panels (C) and (D) compare the estimated neuron densities using the Erö et al. pipeline [8] based on each reference atlas version to literature (bars in brick red with black outline). Each reference atlas version appears in distinct colors: dark blue for CCFbbp, orange for CCFv2, light blue for CCFv3, and green for the manually realigned AV2 and Nissl2.

(A) Focus on the isocortex layer 1/layer 2 boundaries. On the right of each image, layer 1 and layer 2 annotations are shown respectively in blue and orange, on top of the Nissl expression. The density difference between layer 1 and layer 2 creates a visible boundary (see left part of the images). However, in each CCF version, parts of the denser layer 2 are annotated as layer 1, raising the estimated density of neurons for layer 1 (and vice versa with voxels of layer 1 in layer 2). The magnifications shown for CCFv3 highlight some examples of this problem.

(B) Focus on the cerebellar cortex, granular/molecular boundary (Purkinje layer not represented). Each subregion of the cerebellar cortex in the annotation atlas is displayed on top of the Nissl dataset with a distinct color. Molecular (ML) and granular (GL) layers are shown in different shades of the same color, except for CCFv3 which does not distinguish these layers anymore. Here also, the annotation boundaries do not follow the visible density changes between these layers in the Nissl data.

(C) Focus on different regions of the isocortex layer 1. For these regions, the neuron densities produced with the pipeline using each CCF version are greater than their literature counterpart. Using the CCFv3 version (light blue bars) yields the worst estimates. Regions shown: Primary somatosensory area, barrel field (BF); Primary visual area (VA); Primary somatosensory area (SS); Prelimbic area (PV).

(D) Focus on lingula (Ling) and flocculus (Floc) regions of the cerebellum. Results of the manual alignment of the AV2 with Nissl2 are shown with green bars. Density estimates are normalized according to the densities obtained with the CCFbbp. We observe that after manual realignment, estimates tend to be closer to literature values than for the other CCF versions. CCFv3 estimates are not shown, because its AV lacks the separation between molecular and granular layers.

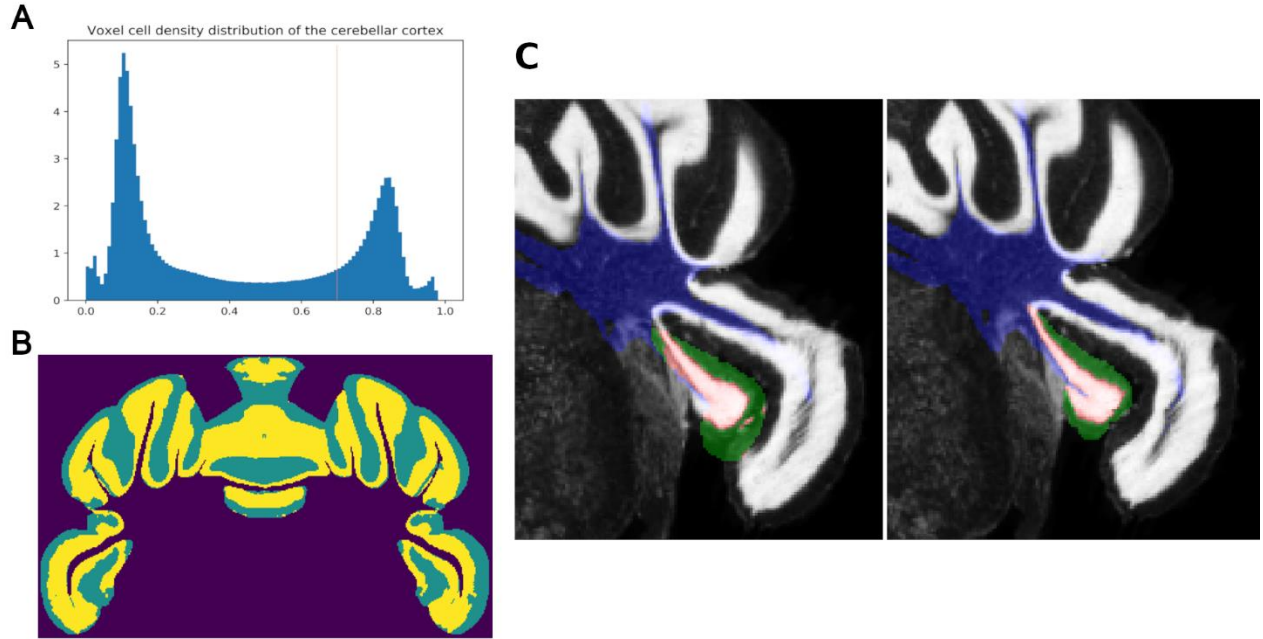


Figure 2-6: **Manual realignment of the cerebellar cortex.**

(A) Normalized histogram of the distribution of Nissl intensity of the cerebellar cortex in the Nissl2 dataset based on the AV2. The two peaks of density correspond respectively to the granular and molecular layer. The vertical orange line corresponds to the threshold chosen to separate the molecular layer from the other layers in the manual correction process.

(B) Coronal slice AV2, showing only the cerebellar cortex annotations after the separation of the layers based on Nissl intensity.

(C) Coronal slice of the cerebellar cortex showing the manual corrections of the annotations of the flocculus. The Nissl volume is shown in levels of gray, white and black representing respectively the highest and lowest Nissl intensity in the slice. Annotations are overlaid in semitransparent colors. The red, green, and blue regions are annotated as respectively granular (+Purkinje) layer, molecular layer, and cerebellar fiber tracts.

We also reproduce the figure 7A from Erö et al. [8], comparing the literature values with their estimates, using the different pairs of reference atlases. However, the results are not conclusive enough to select a reference atlas version because most of the generated cell counts agree with their literature counterparts within their range of variability (see Sup. Figure 1).

## 2.2.4 Orientation and Depth

Most neurons in the brain are oriented towards particular locations in the brain where they send their axons or dendrites. In isocortex, for instance, most neurons point their dendrites towards the pia [4]. Some cell types are also distributed differently according to their position within a region and their distance towards a certain point. This is the case of the stellate and basket cells in the molecular layer of the cerebellar cortex. Basket cells are located next to the Purkinje layer and occupy a third of the volume of the molecular layer while the Stellate cells cover the other two thirds of its volume [14]. Similarly, inhibitory neurons in the hippocampus CAs are more densely located in the upper part of the region [43]. It is therefore critical for neuroscience modelists to be able to compute the vector fields pointing away or towards particular locations in the brain. Moreover, several literature reports depth-based density [43] or connectivity rules [14], that our model would need to integrate. Processing the distance towards particular locations in the mouse brain is also crucial for in-silico modeling. For instance, assigning morphologies to cells requires knowing the space available for them [4].

We want our whole brain models to be used by other researchers in the Blue Brain Project as a scaffold to build circuits of the mouse brain. Our pipeline must therefore provide them the tools to process the orientation fields and distances in the mouse brain.



### 2.2.4.1 Gradient-based approach for orientation field computation

Erö [2] developed a simple algorithm to compute the orientation field within a region, based solely on the annotation volume and a list of source and target region. The algorithm is defined as follows:

1. Create a volumetric dataset where every voxel of the source regions according to the annotation volume is set to -1, the voxels of the target regions set to 1 and the rest is set to 0.
2. Apply a Gaussian filter on the dataset. This step propagates the weight defined at the previous step to the voxels between them.
3. Compute the 3D gradient of the voxels within the region of interest on the resulting dataset and normalize the resulting vector field.

This process can also be performed on a single brain hemisphere which can be useful when the target is situated at the hemisphere border. While being quite simple and fast, this algorithm produces decent results on various brain regions such as the isocortex or the hippocampus CAs. This algorithm has however some drawbacks (see Figure 2-7):

1. The inner order of the region is not guaranteed. For instance, in isocortex, the field will point towards the outside no matter if it means that the vector will cross the regions in the wrong order (arrows on the right of both panels of Figure 2-7).
2. If the target regions are thin enough then the gradient might point towards the interior of the region of interest.
3. The target must be a brain region which sometimes can be surrounding the region. For instance, for the isocortex, the fibers target the pia but the only region in the annotation including it is the outside of the brain. This region surrounds the front of the isocortex which will drive fibers from low layers towards the outside instead of the top part of layer 1 (see Figure 2-7A left arrow).

Overall, this algorithm performs well if the source and target regions are well defined and are separated by the region of interest. Unfortunately, in any other configuration, it suffers from edge effects (see black arrows in Figure 2-7).

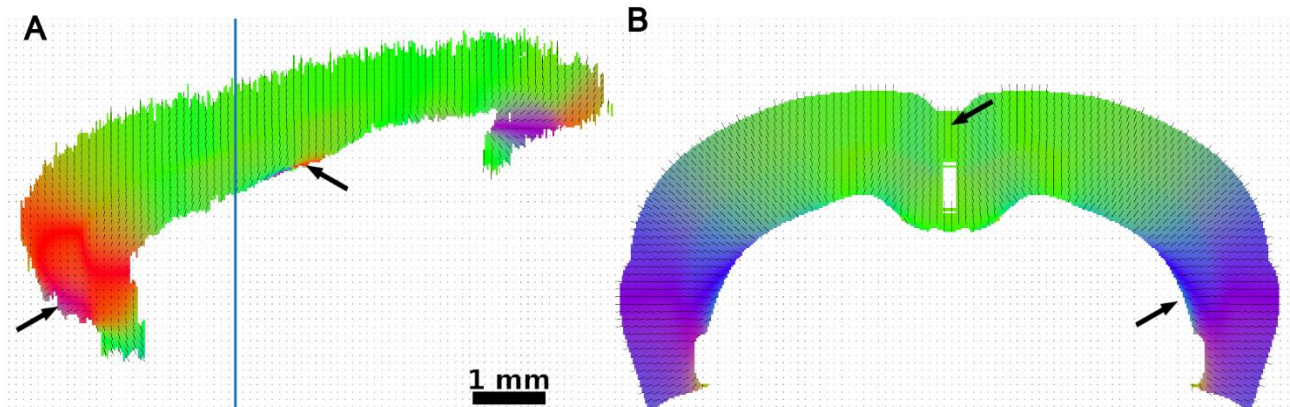


Figure 2-7 **Cortical orientation field using Erö's method.**

Orientation field of the cells in Isocortex. To each voxel of the isocortex  $AV_{BBP}$  is assigned a 3D vector pointing away from the corpus callosum and towards the outside of the brain using Erö's method [2]. The red, green, and blue proportions of the voxel colors correspond to the orientation vectors' norm on respectively the coronal, axial and sagittal plane, black lines are their projected axes. The black arrows point to issues detected in the model.

(A) Sagittal view of the orientation field of the isocortex. The black arrow on the left shows orientation vectors in layer 1 pointing toward the interior of the isocortex. The arrow on the right shows orientation vectors following the same direction as the layer boundaries instead of crossing them. The blue line shows the position of the coronal slice shown in (B).

(B) Coronal view of the isocortex orientation field. The black arrow on the top shows orientation vectors in layer 1 pointing toward the top of the cortex instead of the hemisphere border. The arrow on the bottom right shows orientation vectors following the same direction as the layer boundaries instead of crossing them.

#### 2.2.4.2 Improvement of the gradient-based algorithm

To obtain this orientation field, we created a semi-automated method to generate coordinate systems using only a user-defined list of brain regions as reference. The goal is to compute a vector field pointing towards a general direction, based on a source region and a target region. Within the region of interest, direction vectors are obtained as the normalized gradient of a scalar field. This scalar field is obtained by assigning to every voxel of the brain a user-defined weight representing its distance from the source region (Figure 2-8 AB). For each subregion of interest, a single weight is assigned to every voxel of that subregion and a default value is given to the rest of the brain. For the isocortex, the following weights are assigned to the voxel of the scalar field:

- -2 for the lateral forebrain bundle system (corpus callosum)
- 1 for the layer 6
- 2 for the layer 5
- 3 for the layer 4
- 4 for the layer 3
- 5 for the layer 2 (and 2 / 3)
- 6 for the layer 1
- 0 for the other brain regions

To avoid boundary effects from the outside at the borders of a region, we extended the scalar field of the region of interest to its surrounding voxels (see Figure 2-8B). These surrounding voxels are detected by a shading algorithm which looks for voxels close to annotation borders (i.e., where a change of annotation is occurring). We applied this shading algorithm to each layer of the isocortex, setting their surrounding voxels to the same weight in the scalar field.

An additional scalar shading is computed based on the distance to a subregion of interest identified as a target for fibers. This shading is created to attract the gradient of the voxels in the target region towards the outside. For the isocortex, voxels close to the layer 1 and outside of the brain are assigned to 6 plus their distance to layer 1.

A Gaussian filter is then used to the initialized scalar field and the gradient of the normalized blurred scalar field is eventually returned. The direction vectors are given by this gradient (see Figure 2-8C). This process is applied to all cortical areas at once by defining the white matter as the source region, and the outside of the brain as the target. For the isocortex, we also applied the algorithm on each hemisphere separated from the other as regions close to the middle of the brain are facing each other, which caused the orientation fields to avoid them (see Figure 2-7B left arrow).

I applied the same algorithm to the cerebellar cortex, defining the source region as the cerebellar fiber tracts and the target region as the molecular layer. Each subregion of the cerebellum was dealt separately as several of them are facing each other and can disrupt each other after extending their scalar field. First, we labelled the voxels of the Purkinje layer in each subregion of the cerebellar cortex using our shading algorithm. These voxels correspond to a one voxel-thin layer, at the border between the molecular layer and the granular layer. We changed the voxels of the granular layer; as a manual inspection of the original Nissl experiment images places these neurons within this layer.

The following weights were then assigned to the voxel of the scalar field in these regions:

- -5 for the cerebellar related fiber tracts
- -1 for the granular layer
- 0 for the Purkinje layer
- 1 for the molecular layer
- 3 for the outside of the brain
- 0 for the other brain regions

For each subregion of the cerebellar cortex, we also extended the scalar field to their surrounding voxels and applied an additional shading based on the distance from the molecular layer.

The orientation field algorithm can also be used in the subregions of the Cornu Ammonis (CA), where the source and target regions are defined respectively as the Stratum Radiatum and Stratum Oriens.

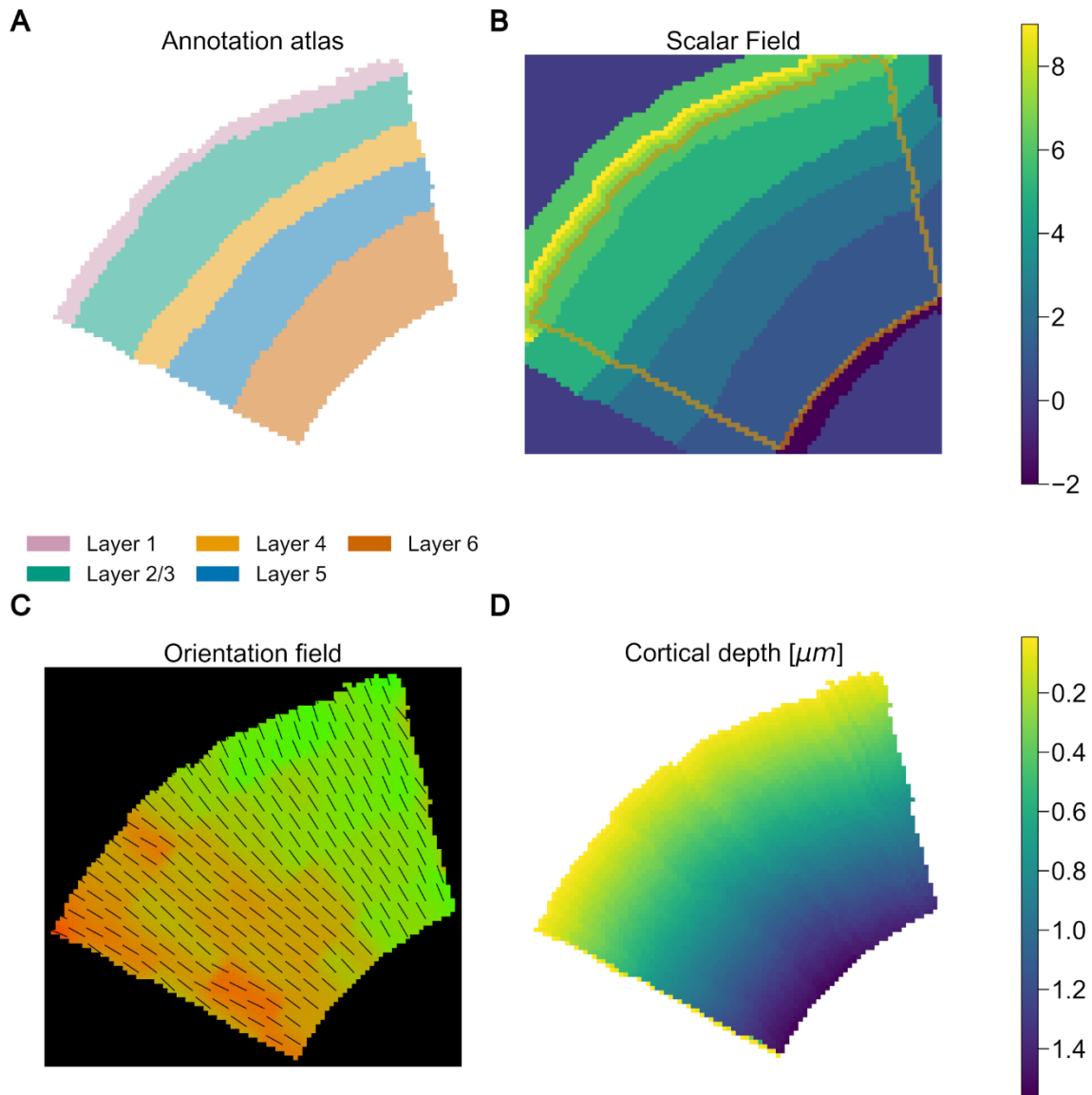


Figure 2-8 **Orientations field and depth computation of the barrel cortex.**

(A) Coronal slice of the AV showing the barrel cortex and its different sublayers.

(B) Coronal slice of the scalar field of the barrel cortex. A weight is assigned to every voxel of the AV. These weights follow the order of crossing the cortex by its fibers from the corpus callosum to layer 1. The borders of the barrel field are highlighted in orange. The weight assigned to the surrounding voxels of the region corresponds to their closest layer's. The weight of the voxels outside the region beyond layer 1 increases as moving away from the barrel cortex.

(C) Coronal slice of the orientation field of the barrel cortex. To each voxel of the AV, a 3D direction normalized vector is computed corresponding to the main axis of the axons in the region. Colors represent the orientation vectors norm on their respective plane, black lines their projected axis.

(D) Coronal slice of the depth according to pia in the barrel cortex expressed in micrometers.

#### 2.2.4.3 An algorithm to compute the distances towards regions of the brain

In Isocortex, the neurons' morphologies are straightly oriented towards the pia. To build a circuit of the Isocortex, a modelist needs to know the distance between each voxel of isocortex and the pia, to fit the correct morphologies. In the cerebellar cortex, however, the neurons' morphology bend according to the curvature of the cerebellum.

To compute the distance for both of these regions, we created an algorithm that reuse the orientation field computed previously. Starting from the position at the center of the voxel, the algorithm adds the unit vector corresponding to the orientation field vector and checks the region id at the new position. This process is repeated until the region at the new position corresponds to the target region. At each step, the unit vector used to move towards the target can be updated according to the orientation field vector at the



new position, if the distance measured has to follow the curvature of the region (case of the cerebellar cortex). Otherwise, the algorithm follows the orientation of the voxel of origin (case of the isocortex).

Within the region of interest, we can additionally record the distance towards each subregion crossed, while moving towards the target (or away from it if we subtract the unit orientation vector). In the Isocortex, we can therefore compute the distance of a voxel from each layer using the same algorithm. This information is useful to know the position of the voxel within a layer. In the cerebellar cortex molecular layer, this can be used to place basket and stellate cells.

## 2.3 Results

### 2.3.1 Effect of Nissl refinement on neuron density estimation.

We applied three different methods to correct for artifacts in the CCFv2 Nissl dataset:

- Smoothing of the Nissl intensity using a Gaussian filter
- Propagation of the distribution of Nissl expression from a reference coronal slice.
- Combination of the two precedent methods (blur + histogram matching).

To evaluate our results, we studied the impact of each method on the estimated neuron density using Erö et al. pipeline [8]. The results can be compared to regional literature findings [41], [42]. We focus first on layer 1 which is particularly sensitive to artifacts in the original Nissl dataset. Indeed, this region has a low cell density, which means that the Nissl intensity in this region should remain very low. However, artifacts tend to increase the Nissl expression (see Figure 2-2 and Figure 2-5 for example). Hence, any correction applied to this region will have a significant impact on our final estimates of cell densities. Figure 2-9AB shows the results of our analysis on isocortex layer 1. We first compared our neuron density estimates in layer 1 in different regions of the isocortex with their equivalent in literature. For most of the regions of the cortex, our corrections were not able to close the gap between our neuron estimates and values from literature (see Figure 2-9A).

The effect of the corrections seems to be different, depending on the position of the region along the rostral-caudal axis. At the front of the cortex, the histogram matching methods tend to raise the neuron densities and lower them in the caudal regions of the isocortex. Figure 2-9BC illustrates this observation. As expected, corrections from histogram matching smoothens the neuron density in the entire isocortex. The histogram matching methods were also able to correct isolated slices with high Nissl intensity due to exposure or artifacts. Smoothing with a Gaussian filter tends to lower the neuron density in layer 1 indicating that local artifacts were indeed raising the neuron densities in this region.

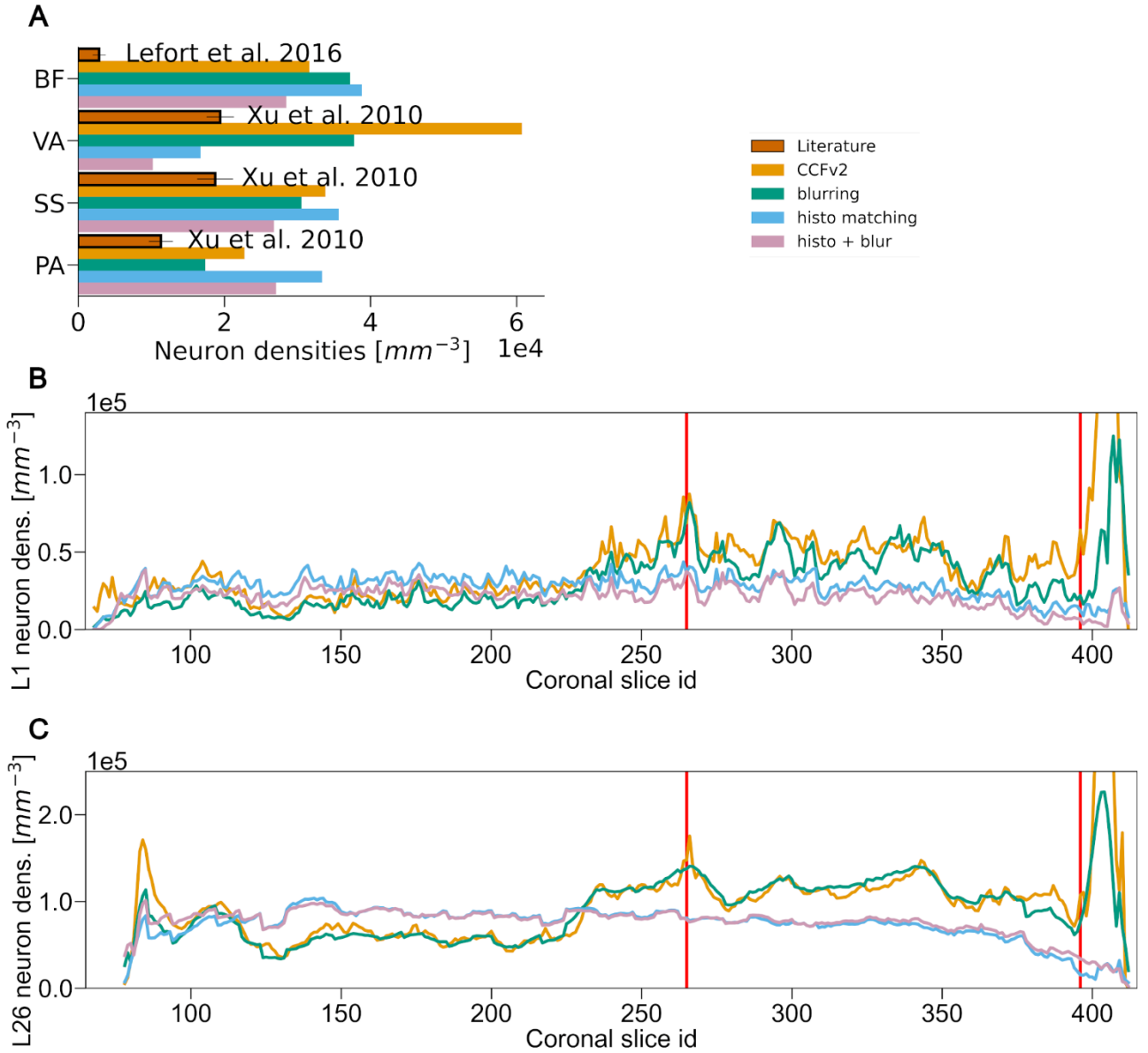


Figure 2-9: **Impact of artifacts corrections on the CCFv2 Nissl volume on cortical neuron densities.**

Analysis of the cortical neuron density estimates from the Erö et al. pipeline [1] after applying the corrections on the CCFv2 Nissl volume. Each CCFv2 Nissl version appears in distinct colors: yellow for the original, green for the dataset after blurring, light blue for the dataset after histogram matching, and purple for the dataset after applying first the blurring and then the histogram matching.

(A) Comparison of the estimated neuron densities, based on each version of the CCFv2 Nissl volume, to literature (bars in red with black outline).

(B) L1 and (C) L2 to L6 neuron density according to the coronal slice id, based on each version of the CCFv2 Nissl volume. Underexposed slice ids were located between slice ids 119 to 240. Overexposed slices were detected at slice 265 and after slice id 395 and are indicated by red lines.

### 2.3.2 Annotation volumes realignment

We realigned the derived annotation volumes AV2' to AV3' using the algorithm described in Section 2.2.2, that leverages the ANTs registration tool [35]. The algorithm follows the AIBS region hierarchy from the root to the most precise regions. We consider each level of the hierarchy completed by the algorithm as a step of the realignment and we evaluate the results of each of these steps to see the progress of realignment. Figure 2-10A shows a sagittal view of the AV2' after the second step of the realignment, i.e., after realigning the brain regions of the two first levels of hierarchy. This step corresponds to realigning the external border of the brain and the border between its main subregions. The sagittal view shows that the border of the brain is smoother and the olfactory bulb shape of the realigned AV2' resembles more closely the shape of its counterpart in the AV3'. However, after several steps of realignment (see Figure 2-10B), the smallest and finest regions of the AV2' are shrunk by the realignment (e.g., L1 and L6b in the isocortex). The displacement field applied to realign a brain region  $r_1$  is limited to the area of its parent region  $r_{parent}$ . If we assume that the

realignment of  $r_{\text{parent}}$  yielded satisfactory results, and if the shape of  $r_1$  in AV2' is strongly affected by the realignment, this might indicate that  $r_1$  borders within  $r_{\text{parent}}$  (or any other child region of  $r_{\text{parent}}$ ) are very different when compared to their counterparts in AV3'. This will induce strong deformations of the voxels within  $r_{\text{parent}}$ . The score metric that we introduced to measure the convergence of the realignment for each brain region (see Equation 2-1) is a good indicator to detect these issues. Indeed, most of the regions, before being realigned start with a score value above 0.5, indicating that at least half of their border voxels are common. However, in the case of the “midbrain, behavioral state related” region, this region starts with a negative score before being realigned by ANTs. The region's borders in AV2' differs drastically from their counterparts in AV3' and the ANTs registration cannot improve the score without strongly deforming the midbrain.

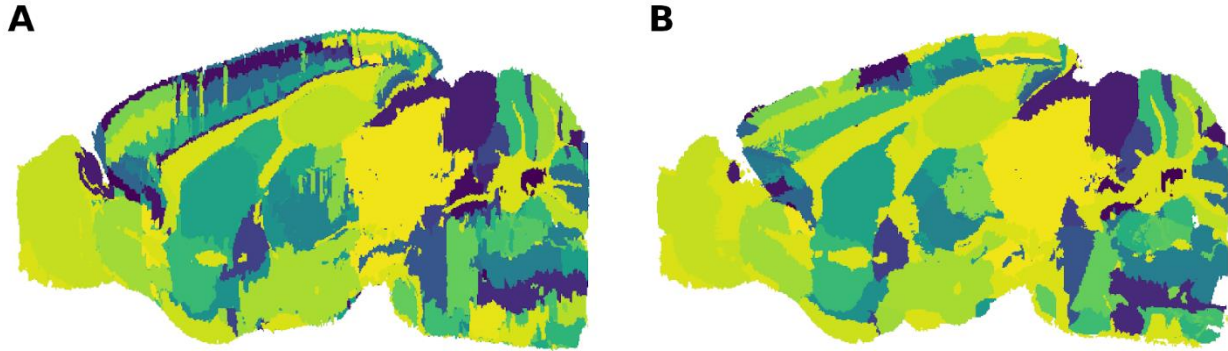


Figure 2-10: **Sagittal views of the merged CCFv2 annotation atlas at different steps of the realignment algorithm.**

To each brain subregion of the annotation volume is assigned a random color.

(A) AV2' after the first two levels of the region hierarchy has been realigned.

(B) AV2' after all region hierarchy levels have been realigned.

To measure the quality of our realignment, we calculated the mean Intersection over Union metric calculated for each pair of coronal slices from the two AVs (IoU, see also Equation 2-2). The results are shown in Table 2-1. As we observed qualitatively in Figure 2-10, the mean IoU of the AV2' over the AV3' improves for the first 2 steps of the realignment before remaining stable or even dropping for the remaining steps. These results indicate that our algorithm can improve the alignment of the AV2 to match the AV3 but require that the merged AVs brain region shapes to be sufficiently similar before applying ANTs realignment tool.

Steps	Origin	1	2	3	4	5	6	7	8	9
Mean IoU (%)	50.14	51.71	53.39	53.05	53.09	54.04	54.85	54.43	54.81	54.76

Table 2-1: **Mean intersection over union of the AV2' annotation atlas at different steps of the realignment algorithm.**

### 2.3.3 Orientation and Depth

We applied our refined algorithm to compute cell orientation and depth in all voxels in the isocortex and cerebellar cortex in the AV<sub>BBP</sub>. The results of the orientation algorithm can be seen in Figure 2-11. Most of the edge effects detected in Figure 2-7, were indeed corrected with this version of the algorithm. In isocortex, the new orientation vectors point towards the closest voxels of the next layer or the pia for L1 instead of directly outside of the brain (see Figure 2-8AB). This improves the frontal part of the region, where the vector fields in some L1 voxels used to point towards the interior of the isocortex (comparing Figure 2-7A with Figure 2-11A). This also creates a visible border between the most frontal part of the cortex, facing the ventral part of the brain, and the middle of the isocortex, which points towards the dorsal part of the brain (see Figure 2-11A). Computing the orientation field for each hemisphere separately helped also with the regions closed to the middle of the brain (comparing Figure 2-7B with Figure 2-11B). We also observe that the orientation field within the regions of interest is less smooth which is expected since we are now using the inner region annotations to drive our orientations.

The algorithm is also accurate in the cerebellum despite the very different and sometimes complex shape of its lobules (see Figure 2-11AC). Since we applied the algorithm on each subregion isolated from the rest of the cerebellum, the orientation fields of lobules facing each other do not impact each other (see Figure 2-11C).

We also computed the depth of each voxel of the two regions which will be used later in this thesis to place the basket and stellate cells in the cerebellar cortex and analyze cell densities distributions in the isocortex.

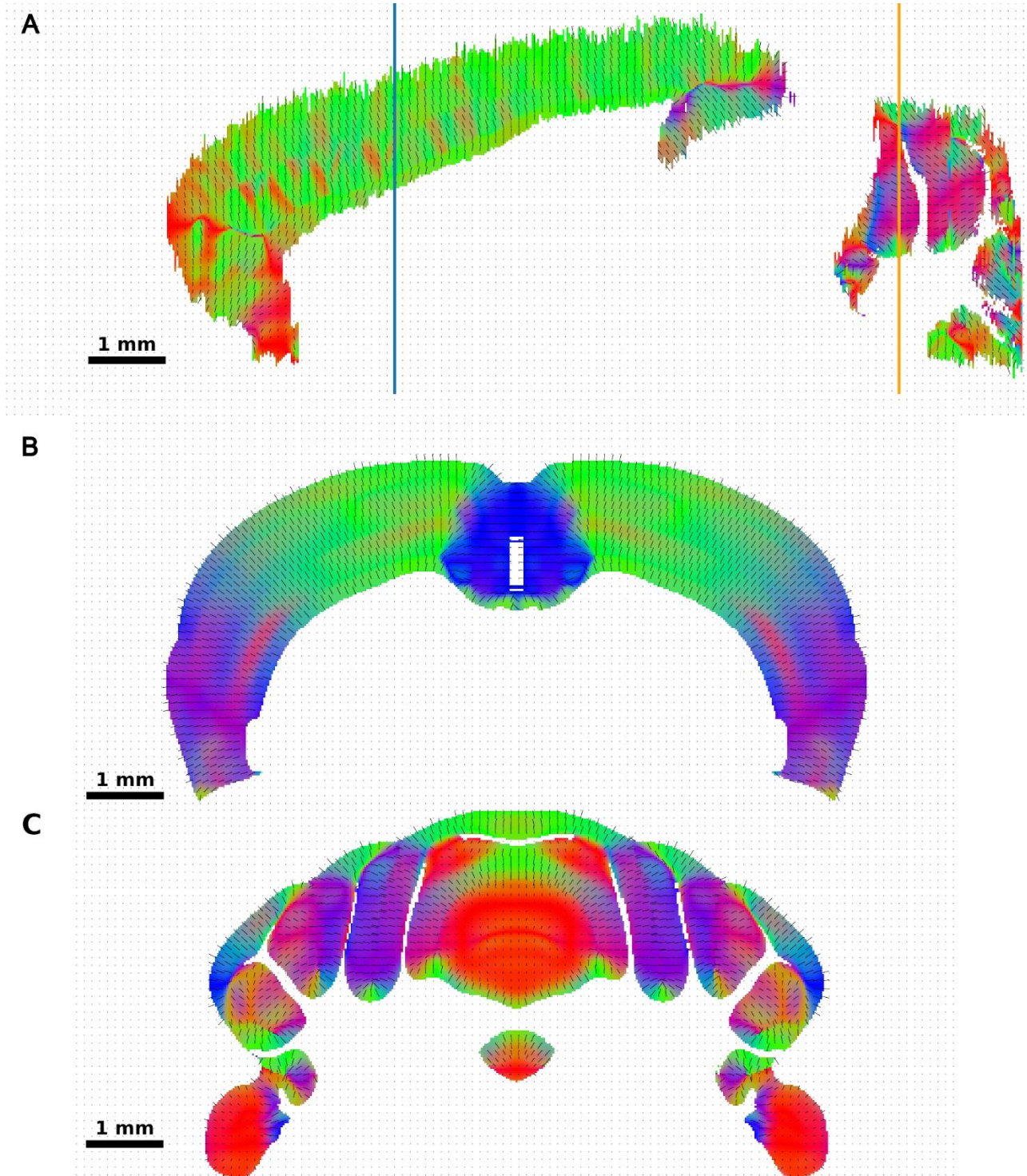


Figure 2-11: **Orientation field of the isocortex and cerebellar cortex.**

(A) Sagittal view of the isocortex and cerebellar cortex orientation fields, based on our new method. To each voxel of the isocortex and cerebellar cortex in the  $AV_{BBP}$  is assigned a 3D vector. The red, green and blue proportions of the voxel colors correspond to the norm of the orientation vectors on respectively the coronal, axial and sagittal plane, black lines are their projected axes. The blue and orange lines in (A) show the position of the coronal slice shown respectively in (B) and (C).

(B) Coronal view of the orientation field of the isocortex.

(C) Coronal view of the orientation field of the cerebellar cortex.

## 2.4 Discussion

### 2.4.1 Nissl2 dataset corrections improve cell density estimations

We have proposed two methods to correct artifacts in a Nissl volume, histogram matching and Gaussian filtering. We applied these corrections on the Isocortex for which we could assume that the Nissl intensity of a layer was similar across all coronal slices. With this assumption, we applied a local Gaussian filter to each layer. We also matched the Nissl intensity distribution to that of a reference slice, using a histogram matching algorithm. This technique should equalize different exposures of the Nissl slices observed in Figure 2-2AD. We tested the effects of these methods and their combination on the cell density estimation pipeline of Erö et al. [8]. We observed that our corrections indeed alleviate most artifacts and yield more accurate neuron density estimates (see Figure 2-9).

However, we could only use these methods to the Isocortex for which we could assume that the Nissl intensity of each layer is homogeneous across all coronal slices. Note, however, that in isocortex, Herculano-Houzel et al. [44] have found that visual areas have higher cell density than the somatosensory areas. Yet, we could not observe such a difference in cell densities in the original Nissl data of the AIBS [7] (see also discussion in Erö [2]). If differences in cell density between different cortical regions were confirmed, we can apply our method to each subregion separately: we could assume that the cell distribution within a leaf brain region is homogeneous and apply our methods to each region. For the histogram matching, we could select a reference slice for each region. Unfortunately, this might not solve the drop of expression observed between distant coronal slices.

The AIBS has sparse Nissl mouse brain slice images that were used as reference to realign the ISH datasets from Lein et al [21] to the AV2. These slices as well as other Nissl datasets could be realigned to Nissl2 to provide an alternative reference on the intensity in certain slices.

### 2.4.2 Realignment of annotation atlases

Our goal is to provide a scaffold model of the mouse brain that can be used as a reference by other researchers as a basis for their models. One requirement to build neural circuits with morphologically detailed neuron models is a smooth AV that resolves all functionally relevant regions. While CCFv3 has the smoothest AV from the AIBS so far, it is incomplete as several important structures are no longer annotated (such as the granular and molecular layers in the cerebellar cortex). Thus, for our purpose of estimating cell densities, we need to augment the available annotation volumes to recover these regions.

To this end, we propose to re-align AV2 to the more recent AV3. The result of this method is a geometric transformation that can be applied to any dataset that is aligned to AV2. The second problem to address is that AV2 and AV3 use different region ids in their annotations. This could be done by producing two derived annotation volumes (AV2' and AV3') with matching region ids. The algorithm realigns the AV region by region and following the hierarchy of brain regions of the AIBS from the main brain region to its leaves.

This technique improved the alignment between the AV2' and the AV3' for the first levels of the brain region hierarchy, but not for the leaf regions that require the higher spatial accuracy. Indeed, despite our efforts to obtain comparable AVs, significant differences remain between the AV2' and AV3' and can sometimes not be corrected by realignment. For instance, the dorsal part of the olfactory bulb in AV3 is touching the frontal part of the isocortex, which is not the case in AV2. When realigning the whole brain volume, our algorithm will therefore struggle to *drag* the boundary of the brain to match AV3. Similarly, borders that exist in one version might not be present in the other. This is the case for the border between the striatum and the hippocampal formation that we see in Figure 2-4C but not in Figure 2-4D. Some AV2 regions are also discontinuous, or very jagged along the rostral-caudal axis. This can make realignment exceedingly difficult for the algorithm. Some regions have also changed shape or position from AV2 to AV3. This is the case for the caudal regions of the isocortex (including the visual areas) which may therefore be impossible to realign without stretching the brain regions drastically. This algorithm therefore requires that the two derived AVs be close in shape and have a similar inner brain region distribution. To produce AV2' and AV3' we tried to find common region ids to replace the ones present uniquely in either AV2 or AV3. During this process, we could also add some constraints to make sure that the overlapping of the brain region is sufficient for realignment. We could also try to smoothen AV2' borders along the sagittal plane and correct discontinuous regions to facilitate the work of the realignment, as was done in Young et al. [45] to improve the annotation volume of the developing mouse.

To recover brain regions that were missing in AV3, alternative solutions can also be found. Brain regions can be split into subregions, knowing the proportion taken by each subregion or using the distance from certain reference points or external regions. For instance, starting from the shape of the AV3 hippocampal CA, we can slice the region following its curvature and knowing the proportion of each layer within. Our depth estimation algorithm could help to process the distance towards the exterior hull of a region. However,



this data might be difficult to find in literature and the result may not correspond to the Nissl volume that we use to estimate cell densities. This means that we would require an extra step to realign the Nissl volume in these regions to the derived AV3.

To integrate experimental data into a common reference space, it is important to develop one unique annotation atlas for the mouse brain. However, on top of the different versions of the AIBS AV, an alternative annotation volume of the mouse brain has been recently produced by Paxinos and Franklin [46]. A solution to merge the AV3 and the Paxinos and Franklin atlases has been proposed by Chon et al. [47] based on the average template brain and cell type marker staining. A future work should investigate if the merged atlas improved on the AV3 and how it can be used in our whole brain model pipeline.

### 2.4.3 Orientation and depth measurements

We described a method to improve Erö's algorithm to compute orientation fields for brain regions. Our version now also takes the various sub-structures of a region into account. It also corrected most of the edge effects that were present in the previous version (i.e., isolated orientation vectors pointing in wrong directions). This used to happen when the target region surrounds the region of interest as in the frontal part of the isocortex. Similarly, holes and misalignment of the coronal slices of the annotation volumes sometimes create a jaggy surface for the boundaries of the inner layers and the border with the outside. Our algorithm extends the region annotations in each direction which compensates for all these effects. We also used a shading algorithm to define the target for the orientation field instead of using the brain region annotations. This way the target is defined according to the last layer inside. This solution was necessary in the cerebellar cortex where the target of the orientation vectors sometimes does not correspond to a brain region (e.g., lobules facing each other, see also Figure 2-11AC).

We computed the orientations field separately for each subregion of the cerebellar cortex. This step was necessary in the cerebellar cortex as the scalar fields, constructed for each subregion, are overlapping. Conversely, in the isocortex, we computed the orientation for the entire isocortex as a whole. We tested the computation of the orientations in the isocortex, treating each subregion individually. This resulted in an even less smooth orientation field (see Sup. Figure 2). The voxels at the border of each subregion are clearly visible in the final field because the target of the orientation field for each subregion is different. This solution might however be preferred over our current solution as it gives a more precise orientation field for each subregion isolated, which might be required for depth computation or morphology placement.

On the other hand, the smoothness of the vector field might be preferred over precision. In this case, Erö's algorithm performs better (see Figure 2-7A and Figure 2-11A). This is due to the underlying annotations which are also very jaggy along the sagittal axis (see Figure 2-1). On top of realigning and smoothening the annotations, we can also smooth the orientation field. For each voxel of the brain, we can indeed override its orientation vector based on the mean direction of the orientation vectors of its neighboring voxels.

We also implemented a new method to compute distances from each voxel according to certain points in the brain. From each voxel of the region, the method follows the orientation field previously computed until it reaches a new voxel and increments the distance traveled. This process is repeated until the program reaches its target. At each voxel crossed, the program can update the direction vector followed to match the curvature of the brain region or remain straight by following the direction vector of the voxel of origin.

This algorithm gives a good approximation of the distance of each voxel according to its target. However, despite the corrections applied by our algorithm to drive the orientations according to the inner layers, some vectors in deeper layers and at the border of the brain region point towards the outside of the brain region. Their distance to the outside of the region is therefore poorly estimated (see left border of the region in Figure 2-8D). We could here again extend the annotations with the shading algorithm to help with the border effect or we could refine the exit condition of our algorithm for the deeper layers.

# Chapter 3 Whole mouse brain cell atlas

This chapter is based on a manuscript that was submitted to PLoS Computational Biology. It describes my work on building a cell atlas of the mouse brain based on heterogeneous datasets. Supplementary analyses have been integrated to provide more details on this project.

The mouse brain contains a rich diversity of inhibitory neuron types that have been characterized by their patterns of gene expression. However, it is still unclear how these cell types are distributed across the mouse brain. We developed a computational method to estimate the densities of different inhibitory neuron types across the mouse brain. Our method allows the unbiased integration of diverse and disparate datasets into one framework to predict inhibitory neuron densities for uncharted brain regions.

## 3.1 Introduction

Erö and colleagues [8] produced the first version of the Blue Brain mouse cell atlas (BBCAv1) which provides a density estimate for various cell (astrocytes, oligodendrocytes, and microglia) and neuron types (glutamatergic and GABAergic neurons) in each brain region defined by the Allen Mouse Brain Atlas [7]. This cell atlas consumes multiple datasets from the Allen Institute for Brain Science (AIBS), including the Allen Reference Nissl volume from Lein et al. [21] and marker expression from in situ hybridization (ISH) experiments to determine the positions of all cells within the annotation volume. Cell densities estimates are finally placed within the AIBS annotation volume. The authors used global constraints from the literature [17], such as the total number of cells and neurons in the mouse brain from Herculano-Houzel et al. [22] to avoid any bias toward a particular region.

Erö et al. argued that literature values are often inconsistent and are therefore unreliable [8], [17]. Hence, the authors used solely whole-brain values from literature to constrain their estimates of cell densities in every brain region. However, there are also several regions for which a consensus on cell composition has been reached (e.g., for the cerebellar Purkinje layer, see review from Keller et al. [17]). For instance, in all regions of the isocortex, it is well-known that layer 1 regions should only be composed of inhibitory neurons [13]. To take this aspect into account, the BBCAv1 had to manually integrate those regional constraints a posteriori (e.g., force each neuron in isocortex layer 1 to be inhibitory) which in turn contradicted the initial global value used in the pipeline. Moreover, the BBCAv1 strategy yields regional estimates that sometimes do not match region-specific estimates from the literature (e.g., cell estimates for cortical subregions compared to Herculano-Houzel et al. [44]). Finally, to obtain the composition with respect to an even more fine-grained classification, using the strategy of BBCAv1, we require brain-wide estimates for every new cell type. These constraints may not yet exist.

The BBCAv1 was further extended by Erö [2] to refine the neuron densities generated with the pipeline with specific brain region neuron type distributions from literature [4], [5], [14], [19], [48]. These types are abstract and can correspond to morphological or transcriptomic types. Then a distribution of electrical types is further assigned to each neuron type. These electrical types define the neuron's electrical behavior which can be further modeled by a point neuron model with a specific set of parameters. These regional neuron/electrical type distributions override the previously assigned excitatory and inhibitory labels from the BBCAv1. The whole brain distribution from the Cell atlas was indeed considered less accurate as it is based on whole brain datasets and global estimates. For example, the cellular composition of the somatosensory cortex, lower limb from Markram et al. [4] was applied to the entire isocortex. Moreover, the same cell type and electrical type proportions were used across all cortical brain regions. This creates a uniform distribution of cell types in the cortex, which is thus only constrained by the neuron distribution of the cell atlas.

In this chapter, we will therefore devise a method to improve on the BBCAv1, extending it to include more inhibitory types. In particular, we want this updated version of the cell atlas model to be capable of integrating various sources of literature on cellular density into a consistent cell atlas. We also need to define a method to link the neuron and electrical types from literature to the general types of the BBCAv1 rather than overriding them to ensure consistency between the different versions of the cell atlas.

We extended the workflow from Erö et al. [8], using estimates of cell densities for individual brain regions from the literature instead of whole-brain (global) estimates for specific cell types. Specifically, using additional ISH experiment datasets from the AIBS, we were able to further refine the GABAergic neuron densities from Erö et al. [8] and estimate the densities of GABAergic subclasses. This

includes parvalbumin (PV+), somatostatin (SST+), vasoactive intestinal peptide (VIP+) immunoreactive neurons and the remaining inhibitory neurons.

PV+ neurons are fast spiking, resulting in fast and efficient suppression of their surrounding neurons' activity [49], [50]. A deficit of PV+ neurons seem to be linked to neural diseases such as schizophrenia [51], Alzheimer's [52] and autism [53], [54]. SST+ neurons participate in producing long range inhibition [50]. This population of neurons decreases with age and seems to contribute to the development of Alzheimer's disease [55]–[57]. VIP+ neurons, on the other hand, tend to target other inhibitory neurons in the isocortex, playing the role of local disinhibitors [58]. SST+ and VIP+ neurons are also involved in the circadian clock which regulates daily brain activity [59].

We also highlighted the fact that the alignment of multiple datasets to a common coordinate system is essential to obtain the best results in terms of cell density estimations. We therefore used a tool with deep learning techniques to automatically realign ISH datasets to our reference system. This tool replaces the manual realignment step in Erö et al.'s pipeline. We combined all these new tools and methods to refine Erö's workflow [8] and produce a second version of the Blue Brain Mouse Cell Atlas pipeline (BBCAv2).

Our whole brain model pipeline needs to further extend the cell atlas molecular identities with morphological and electrical types (me-types). This problem is not trivial as molecular identity often does not directly translate to morphological and electrical features [60]. However, many studies provide information about circuit composition through this me- classification [4], [61], [62]. Previous approach from Erö [2] was overriding the BBCAv1 whole brain densities with regional distribution from literature which were considered more precise.

We present in this chapter another approach to map the neuron molecular identities to morphological and electrical types. We used Roussel et al. technique [63] to map well identified cortical me-types from Markram et al. [4] to our GABAergic neuron subclasses for all regions of the isocortex. This method has the advantage to maintain the whole brain density distribution obtained at the end of the BBCAv2 pipeline

These refinements to the model will help neuroscientists to get a better understanding at cellular composition in the brain and will pave the way for more accurate in silico reconstructions of brain tissues.

## 3.2 Methods

### 3.2.1 Overview of the pipeline

#### 3.2.1.1 *Original Pipeline*

In this chapter, we refined several steps of the BBCAv1 to facilitate the integration of more and more specific cell types without having to rely on the total number or ratio of these cell types. We started with the original Blue Brain Mouse Cell Atlas workflow (BBCAv1) from Erö et al. which consisted of six steps [8]:

1. The Nissl-stained brain slices from Dong [7] and the corresponding mouse brain Annotation Volume (AV) were manually realigned along the sagittal axis. Every ISH dataset [21] used in the following steps was also manually realigned to the Nissl volume.
2. The annotated Nissl volume from the AIBS was combined with a total number of cells in the mouse brain from Herculano-Houzel et al. [22] to estimate the number of cells in each brain region.
3. A combination of genetic marker datasets and a global ratio of glial cells in the brain from Herculano-Houzel et al. [22] was used to distinguish neurons from glial cells.
4. Glial cells were labeled as astrocytes, oligodendrocytes or microglia based on whole brain ratios together with their respective ISH datasets.
5. GAD67 (glutamic acid decarboxylase) marker experiment from Lein et al. [21], associated with a global number of inhibitory neurons from Kim et al. [16], was used to distinguish the excitatory from the inhibitory neurons. This number stands for the sum of PV+, SST+ and VIP+ reacting cells in the brain and was assumed in this pipeline to represent the entire inhibitory population.
6. Manual correction of purely inhibitory regions: molecular layer regions of the cerebellum, layer 1 of the isocortex, the reticular nucleus of the thalamus, and the striatum were considered to be fully inhibitory, which means that all their neurons were labeled as GABAergic.



### 3.2.1.2 Adjustments for improved pipeline

Step 1 of the BBCAv1 involved manual selection of landmark points on each slice of the AV, and ISH data to realign to the Nissl volume. This step was not only labor intensive but also error prone. In our updated workflow, we replaced the manual realignment by a novel deep-learning based alignment algorithm by Krepl et al. [36].

In step 5, the total number of inhibitory neurons in the mouse brain was a global constraint. Its biological value was taken from Kim et al. [16] and represented the sum of SST+, PV+, and VIP+ neurons. However, these three types do not represent all inhibitory neurons (e.g., LAMP5 cells in the isocortex [60]). This means that the total number of inhibitory cells in BBCAv1 was underestimated, despite the corrections applied in step 6.

We therefore reworked steps 5 and 6 by integrating more literature estimates of neuron type densities as well as by using additional ISH datasets. This not only improved our estimates of the excitatory to inhibitory ratios in the brain, but also provided more precise estimates of GABAergic neuron types, including PV+, SST+, and VIP+ cells. However, as we will discuss later, the spatial resolution (200  $\mu\text{m}$  between each coronal slice) and the whole-brain coverage of these datasets are not sufficient to estimate the density of all inhibitory subtypes in each voxel of the mouse brain, as it was done in BBCAv1. Therefore, we computed the mean density of each inhibitory type for each region of the AV, instead of estimating the density per voxel.

Since several AV and Nissl volumes were recently released [39], we need to choose the best combination to estimate cell densities (see Section 2.2.3). Our earlier observations indicated that misalignments lead to overestimation of densities for low density regions and to underestimation in high density regions, reducing the differences between regions and thus, making their values more homogeneous. We quantify this by calculating the standard deviation of the neuron count distribution for each CCF. We find that the standard deviation is the largest for the version of CCFbbp (Erö et al.  $2.82 \times 10^6$ , CCFv2  $2.26 \times 10^6$ , CCFv3  $2.57 \times 10^5$ ) and therefore least affected by misalignment. We therefore adopt CCFbbp throughout the rest of this thesis.

### 3.2.1.3 New pipeline

Figure 3-1 shows the new pipeline for the Blue Brain Mouse Cell Atlas version 2 (BBCAv2). It consumes data from the AIBS in the form of image stacks of stained coronal brain slices, as well as a combination of annotation and Nissl reference atlases.

The BBCAv2 has four main steps:

1. In the first step, the different image stacks from ISH experiments from the AIBS are automatically aligned and registered to a reference volume. We describe this step in Section 3.2.2.
2. The second step generates cell, neuron, and glial densities for each region of the AV. It corresponds to steps 2 and 3 of the BBCAv1 pipeline.
3. We devise assumptions to estimate densities of different inhibitory neuron types in a coherent framework (see Section 3.2.3) and apply them for step 3 and 4. In step 3, we assume a correlation between genetic marker expression and cell type density gathered from literature (see Section 3.2.4) and the previously realigned and filtered ISH datasets. The details of this step are described in Section 3.2.5.
4. Step 4 integrates the results of step 3 into a consistent estimate of the cell densities for each inhibitory subtype, according to our assumptions (see Section 3.2.6). Finally, the cells are placed in the brain volume to produce the updated BBCAv2 model (see Section 3.2.7 and 3.2.8).

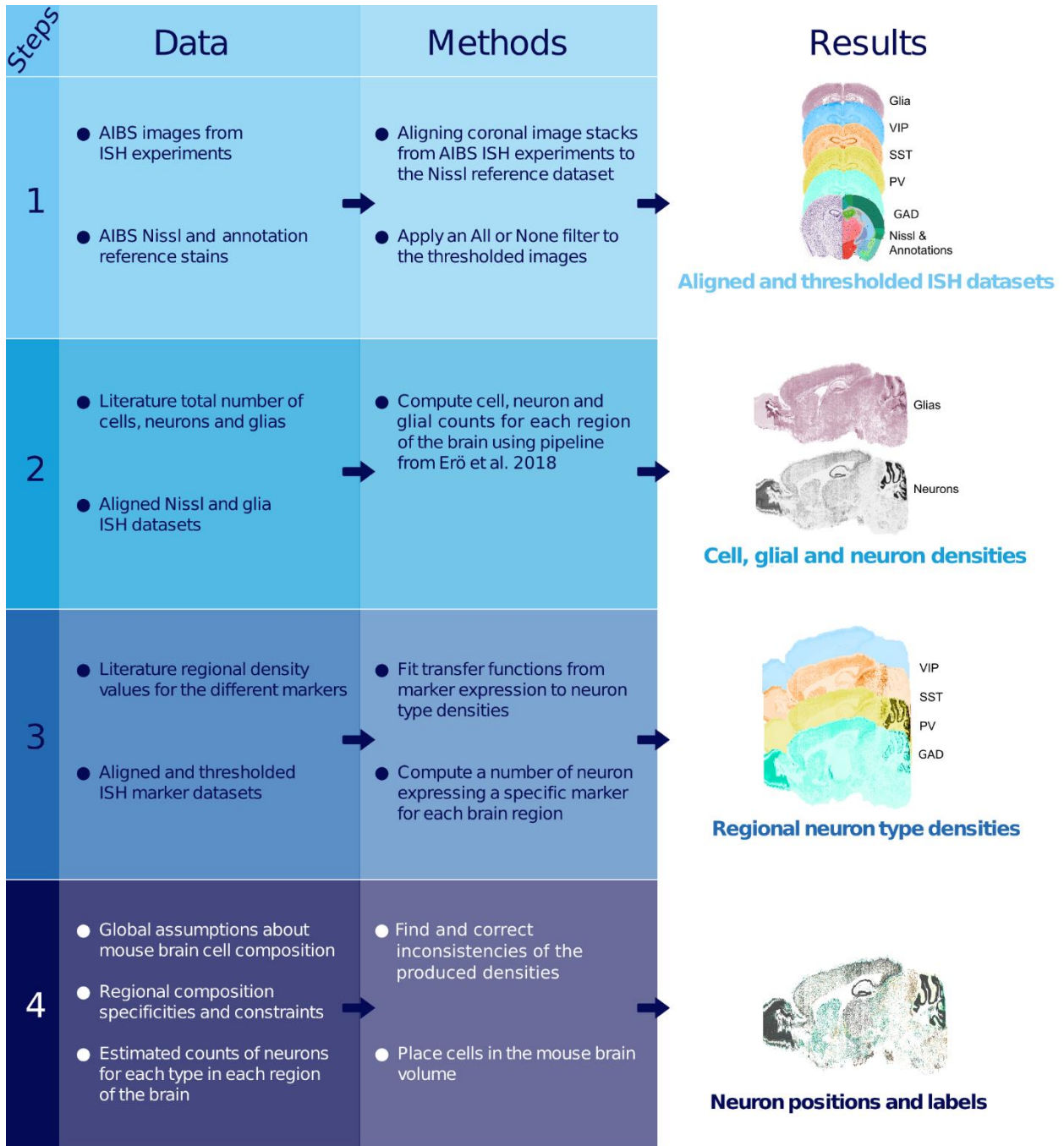


Figure 3-1 **New workflow for the Blue Brain Mouse Cell Atlas (BBCAv2).**

Steps of the pipeline are displayed in rows with the data consumed in the left column, the method used in the middle column, and the produced results on the right. Each step of this pipeline builds on the results of the previous steps.

### 3.2.2 Alignment of in situ hybridization coronal brain slices

For step 3 of the BBCAv2, we consume ISH datasets from the AIBS [21] which are sectioned in the coronal plane (one 25  $\mu\text{m}$  thick section every 200  $\mu\text{m}$ , see Figure 3-2D). However, the individual images are not aligned to the same Nissl volume that we use to estimate neuron densities. In the BBCAv1, ISH images slices were manually realigned to the Nissl volume with a non-rigid landmark-based algorithm [8]. This task was very tedious and time consuming because it required an expert to manually choose common landmarks between adjacent sections. For our new pipeline, we therefore replace this step with a fully automatic machine learning based method.

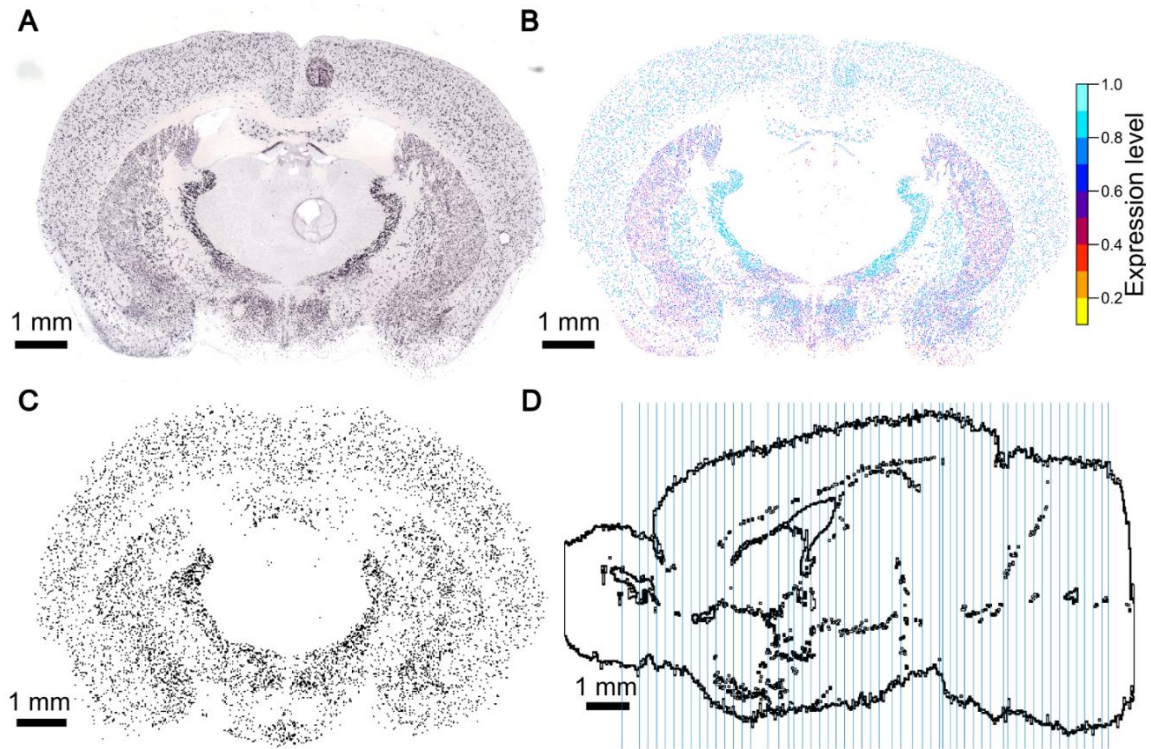


Figure 3-2 **AIBS GAD67 ISH experiment data.**

(A) raw image, (B) filtered image and (C) realigned binarized image. This ISH experiment highlights inhibitory cells of the mouse brain. The raw slice image (A) contains artifacts and a minimum expression value offsets the dataset. The filtered image (B) is derived from the raw image (A) by the AIBS. Somas reacting to GAD67 are detected in the raw image and isolated from the background. The distinct colors of the cells in (B) represent the distinct levels of expression [64]. The filtered images are realigned and thresholded to obtain the binarized image shown in (C). (D) Positions of the coronal stained slices, shown in blue on a sagittal slice of the CCFbbp brain volume

The different steps of the automatic realignment of every ISH dataset from the AIBS to the reference atlas are as follows:

- Original ISH images (or raw images - see Figure 3-2A) and their thresholded versions, later called filtered images (see Figure 3-2B), are downloaded from the AIBS website. The filtered images isolate the somas reacting to their specific marker from the brain background in the raw image.
- Each image is scaled, positioned, and rotated according to the reference brain volume, using the provided metadata. The metadata includes the position of the corners of the images because the slices are not perfectly vertical. Here we assumed the slices to be vertical as the algorithm was trained with this assumption and it makes the interpolation between slices easier. Hence, we used the mean position of the coronal ISH images in the AIBS average brain along the rostral-caudal axis also from the metadata. The AV2 and Nissl2 volumes have been realigned to the average brain, so the Nissl coronal slices should correspond to the ISH's. For the CCFbbp, used here, for each ISH coronal slice to realign, we manually selected the best corresponding reference Nissl slice.
- For each dataset, every coronal section is automatically realigned to the corresponding anatomical section of the Nissl-stained mouse brain using the Krepl algorithm [36]. Here, the raw in situ images are used as they provide more landmarks for the algorithm.
- Resulting displacement fields are then applied to the filtered images of the corresponding ISH experiment.
- A 3D volume for each gene is created by down sampling the images to match the final voxel dimensions of the reference atlas. Missing sections between the 2D coronal slices are generated with a linear intensity-based interpolation.

We download and realign the following datasets from the AIBS ISH portal [65]:

- Parvalbumin: Experiment id #868
- Somatostatin: #1001
- Vasoactive Intestinal Peptide: #77371835
- GAD1: #479 (equivalent of GAD67 see Figure 3-2) and #79556706 (used in Section 3.3.5)

The resulting 3D voxel-based datasets are consumed by our pipeline to estimate inhibitory neuron subtype densities (see Section 3.2.5).

### 3.2.3 Assumptions underlying the estimation of inhibitory subtypes densities

We wish to obtain the number of inhibitory neurons and their subtypes for each region of the AV1. The regions are indexed by region ids  $\mathbf{R} = \{0, r_1, r_2, r_3, \dots, r_n\} \in \mathbb{N}^{861}$ , with 0 denoting the outside of the mouse brain. We are particularly interested in inhibitory subtypes expressing the markers GAD67, PV, SST and VIP.

First, the number of neurons can be expressed as the sum of the different subtypes in each region  $r$ ,  $\forall r \in \mathbf{R}$ :

$$nNeu_r = nInh_r + nExc_r + nOther_r \quad \text{Equation 3-1}$$

where  $nNeu$ ,  $nInh$ , and  $nExc$  are respectively the total, the inhibitory, and the excitatory neuron numbers. As reported by Zeisel et al. [66], these populations seem to be mutually exclusive. The term  $nOther$  represents the number of neurons that are neither excitatory nor inhibitory, meaning they could be purely modulatory neurons.

The number  $nInh$  can be estimated by the number of neurons, reacting to the markers GAD67 and GAD65 [66] which implies that  $nInh = nGAD$ . We further subdivide  $nGAD$  into inhibitory subtypes positively reacting to the markers PV, SST and VIP. This yields the following sum,  $\forall r \in \mathbf{R}$ :

$$nGAD_r = nPV_r + nSST_r + nVIP_r + nRest_r \quad \text{Equation 3-2}$$

Here  $nPV$ ,  $nSST$  and  $nVIP$  are respectively the number of PV+, SST+ and VIP+ immunoreactive neurons in the brain region  $r$ . The term  $nRest$  corresponds to the number of inhibitory neurons which react neither to PV, SST nor to VIP (InhR neurons), including, for instance, the LAMP5 cells in the isocortex [60]. InhR neurons might define various populations of GABAergic cells, reacting to a wide variety of markers [66], which makes it more difficult to estimate  $nRest$ . It is therefore easier to estimate  $nGAD$  from literature and the marker GAD67 (for regions not covered by literature). Similarly,  $nPV$ ,  $nSST$  and  $nVIP$  can be estimated from literature and the markers PV, SST and VIP, respectively. Finally, we can estimate  $nRest$  by subtraction in Equation 3-2. The remaining neuronal populations ( $nExc + nOther$ ) can be deduced from  $nGAD$  (Equation 3-1), and the neuron distribution ( $nNeu$ ) obtained by step 2 of our BBCAv2 pipeline (see Figure 3-1).

In this estimation, we rely on the following assumptions:

1. Every GABAergic neuron expresses GAD67 and every GAD67 reacting cell is a GABAergic neuron. This genetic marker is indeed responsible for over 90% of the synthesis of GABA [67], [68]. Additionally, no cells expressing GAD65 without expressing GAD67 have been reported in the RNA-sequence study from Zeisel et al. [66].
2. GAD67, PV, SST and VIP are only expressed in neurons [59], [66], [69], [70].
3. PV+, SST+ and VIP+ populations are non-overlapping i.e., there are no cells in the mouse brain that co-express a combination of these markers. This assumption is supported in the isocortex and hippocampus by transcriptomic studies such as Huang and Paul [60] and Zeisel et al. [66] and we extrapolate these findings to other areas.
4. Every PV+, SST+ and VIP+ neuron also expresses GAD67 as observed by Celio [71] and Tasic et al. [72].
5. Neuronal composition is homogeneous within subregions at the lowest level of the AIBS region hierarchy.

We will also consider the cell, glia and neuron density distributions obtained in step 2 of the pipeline (see Figure 3-1) to be correct. These numbers have been validated against literature by Erö et al. [8] and can be refined in the future.

Based on these assumptions, BBCAv2 will provide estimates of the densities of GAD67+, PV+, SST+, and VIP+ neurons for each region of the brain. In contrast with BBCAv1, this new version will present estimates that are as close as possible to available literature values.

### 3.2.4 Literature review of estimates of the mouse inhibitory neuron densities

For the wild-type adult mouse, the AV1a defines 861 regions. Ideally, we would need at least one literature value for the cell density of each marker in each of these regions. We should also have multiple estimates of the same quantity to increase the reliability of the estimation. The extensive review of Keller et al. [17] provided us with references for inhibitory neuron numbers in the mouse brain (e.g., [51], [73]–[76]). Among them, the extensive dataset of Kim et al. [16] gives a complete measurement of the distribution of the PV+, SST+ and VIP+ cells in the mouse brain, covering 97% of the regions of the AV1a. We also made a comprehensive literature search for cell densities or cell counts of GAD67, PV, SST, and VIP immunoreactive neurons in the mouse brain (wild-type species, see supplementary methods for more details). In total, we found 32 references providing values for GAD67, 29 for PV, 14 for SST and 9 for VIP (see Table 3-1). These estimates are in the form of density of neurons or cell counts within either a volume or a slice. For counts with no volume provided, we used the corresponding volume within the AV to estimate a mean density within the respective region. Finally, some literature sources report percentages of labeled neurons or cells within a region volume. For these estimates, we used the distributions obtained in step 2 of our pipeline (see Figure 3-1).

	Number of papers	Number of regions covered (out of 861)	Coefficient of variation	
			Median CV	CV Skewness
GAD67	32	151 (17.5%)	0.20	1.19
PV	29	836 (97.1%)	0.24	2.94
SST	14	849 (98.6%)	0.13	3.74
VIP	9	849 (98.6%)	0.22	1.90

Table 3-1 **Summary of papers reviewed sorted by marker.**

The number of papers from which we extract density values as input for our model are listed in the first column. The table also shows how many regions of the brain are covered by these values. For each region where we obtain an estimate from at least 2 different papers, we can compute a coefficient of variation. From the list of these coefficients of variation, we then extract the median values and the skewness of the distribution, defined as  $s = E \left[ \left( \frac{CV - \text{mean}(CV)}{\text{std}(CV)} \right)^3 \right]$ , which are displayed for each marker in the last two columns.

Erö et al. [8] considered the neurons of some regions to be inhibitory only (step 6 of the BBCAv1 workflow). Similarly, we imposed this rule on layer 1 of the isocortex, the molecular layer of the cerebellar cortex and the reticular nucleus of the thalamus. This means that for these regions, we assume that the density of GAD67+ neurons correspond to its neuron density. Thus, in total we collected GAD67+ density estimates for about 17.5% of all regions of the AV (see Figure 3-3A).

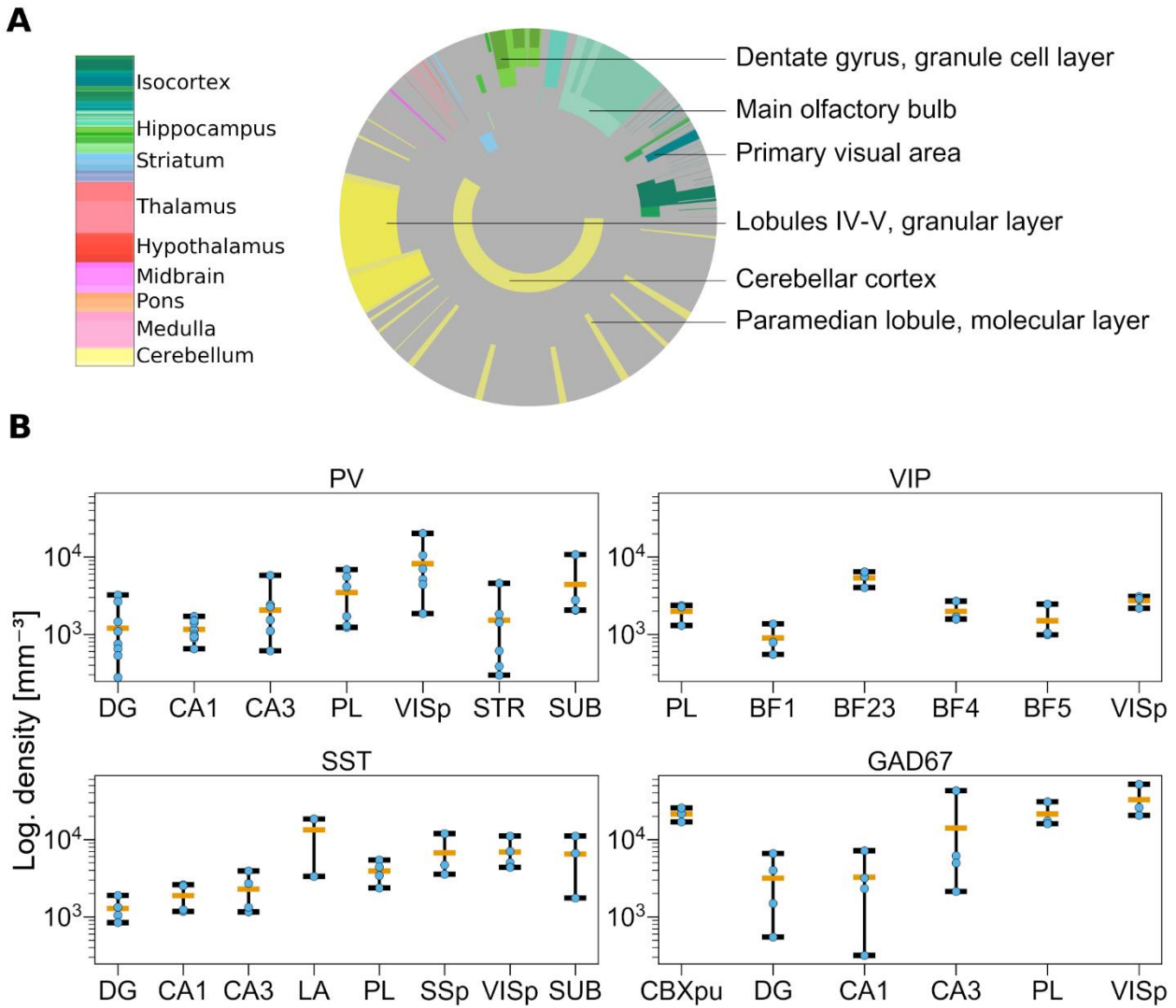


Figure 3-3 **Overview of published data on inhibitory neuron densities in the mouse brain.**

(A) Illustration of regions for which information on densities or absolute numbers of GAD67+ neurons were published. The disk is divided into rings and sectors. Each ring represents a hierarchical level in the AV and sectors represent the contained regions. The center of the disk represents the entire brain, each surrounding ring then represents the next hierarchy level. Colored areas represent regions where at least one study reports cell numbers or densities. The size of the sectors corresponds to the number of neurons relative to the number of neurons in the brain. The colors correspond to regions according to the AV such as the cerebellar cortex in yellow and cortical areas in green.

(B) Variability of published cell density estimates. Each panel shows the distribution of reported cell densities (log scale). For each region, data points are depicted in blue, the average in orange. The minimum and maximum values are indicated by the whiskers. Regions shown: Dentate Gyrus (DG); Field CA1 (CA1); Field CA3 (CA3); Prelimbic area (PL); Primary Somatosensory area (SSp); Primary Visual area (VISp); Striatum (STR); Subiculum (SUB); Primary somatosensory area, barrel field (BF); Lateral Amygdalar nucleus (LA); Cerebellar cortex, Purkinje layer (CBXpu).

For some regions such as the CA1 field of the hippocampus, we observe a large variability of the literature estimates (see Figure 3-3B): spanning from 317 GAD67+ cells per mm<sup>3</sup> computed from Han et al. [77] to 7166 GAD67+ cells per mm<sup>3</sup> according to Jinno et al. [43]. In addition to the inter-subject variability, these large variations can arise from a combination of distinct factors, such as differences in the staining methods, the counting methods, or the methods to determine subregion boundaries (annotations) [15], [16].

Our literature review highlights the current lack of knowledge and consensus on cell composition in the mouse brain, emphasized by the large variability of published data. In the following section, we will describe a method to estimate cell densities in regions for which no literature values were published.



### 3.2.5 Transfer functions from marker expression to cell density

We use the realigned ISH datasets obtained in Section 3.2.2 to estimate the densities of inhibitory cell types in regions where no literature values were reported.

Different cell types express genetic markers at distinct levels [66]. Nonetheless, we want to count every cell reacting to the markers. We do this by applying an all-or-none filter to the realigned filtered image stacks (Figure 3-2B) using Otsu's method [78] to define the threshold. This forces each pixel showing a cell reacting to the marker to the value 1 and all other pixels to 0 (binarized image stacks see Figure 3-2C). The more pixels in a region are set to 1, the more densely the region is filled with marker-positive neurons. We therefore expect, for a specific brain region, that its mean pixel intensity correlates with the density of marker positive cells. We use this relationship to construct transfer functions linking densities of GAD67+, PV+, SST+, and VIP+ cells reported in the literature to the mean intensity in each region of the AV (see Figure 3-4).

We leverage every literature density value available for each subregion of the brain, assigning the same weight to each point. We compare separately the cerebellum, isocortex and the rest of the brain. We choose this division because in the cerebellum, cell densities are very high compared to the rest of the brain [22]. And inhibitory cell densities and cell type composition across layers of the isocortex are similar in all of its subregions [16]. The results of the comparison between region mean pixel intensity and cell type density are shown in Figure 3-4.

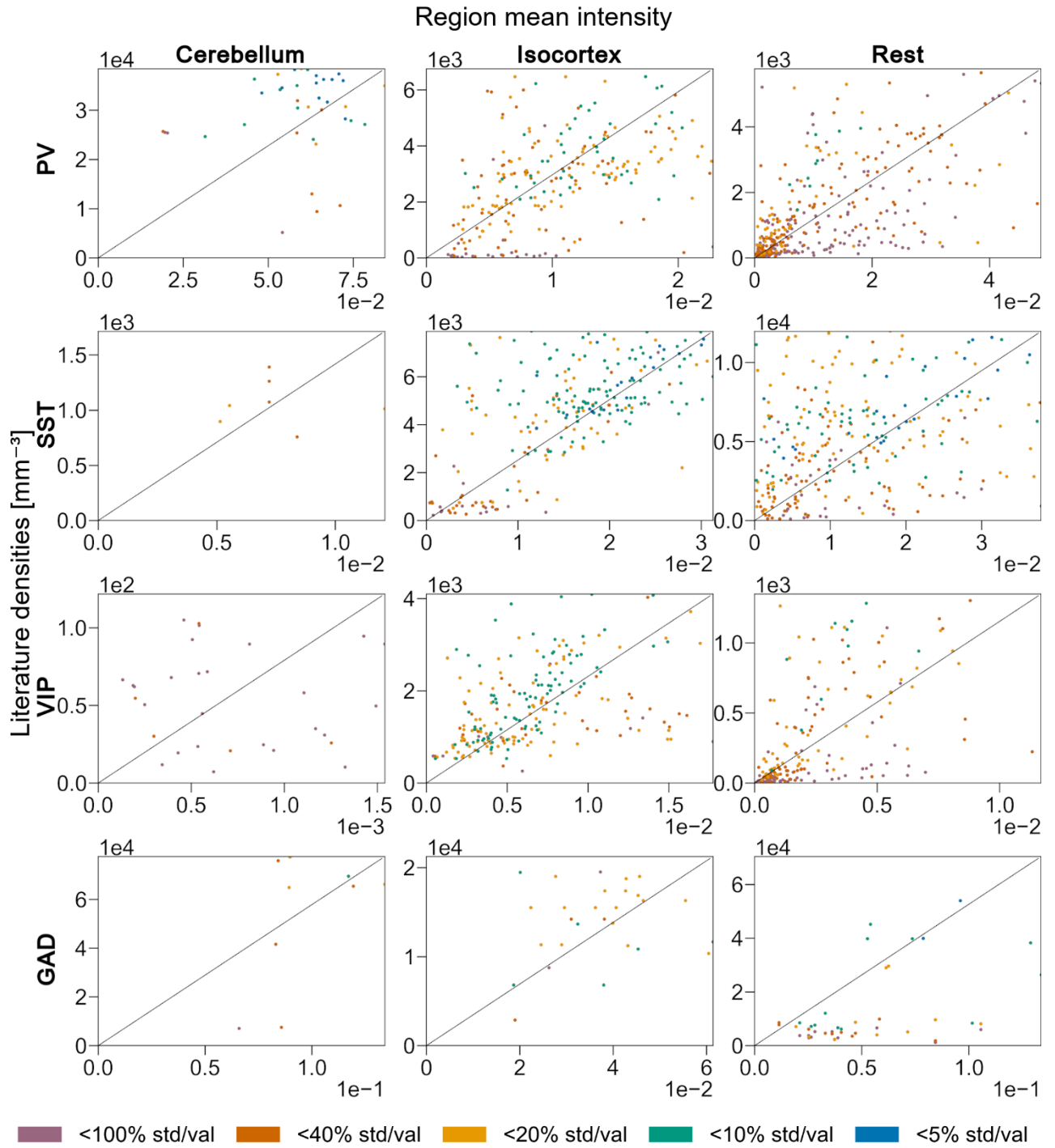
We expect a region with no inhibitory neurons to have a null mean intensity and conversely a region with null mean intensity should not host any inhibitory neurons. However, we do not observe this property in many cases. This observation can be linked, on the one hand, to the misalignment between the annotation atlas and the ISH datasets (cells reacting appear in the wrong region) or on the other hand, to the insufficient coverage of the brain by the ISH data (region of interest does not appear on the ISH slices, see Figure 3-2D). Therefore, these points are excluded from Figure 3-4. Additionally, as we discussed in Section 2.2.3, the density estimation in fully inhibitory regions tends to be highly impacted by misalignment. The points linked to these regions are therefore also excluded from the figure.

At the first glance, there seems to be no clear relationship. However, as we have seen above there is a large variability in the cell densities reported by literature (see Figure 3-3). This explains the spreading of the points along the y axis in the different panels of Figure 3-4 (different point colors). Similarly, points along the x axis are also spread out because of the sparseness of the ISH datasets (see Figure 3-2D) and its misalignment to the AV (see Figure 2-5). This variability might also be linked to the different cell sizes which impacts the region mean intensity. These two sources of error make it difficult to estimate a transfer function that maps region mean intensity to cell type density. Logically, we expect that there is a monotonically increasing relationship between region mean intensity and cell type density — one for each marker and each main division of the brain. The simplest assumption is that these relations are linear and can be calculated by fitting (black diagonals in Figure 3-4).

To evaluate the quality of the fit from region mean intensity to densities of cells, we compute the coefficient of determination  $R^2$  for each linear function  $f$ , which fits the cloud of points  $(x, y)$  where  $x$  is the region mean intensity and  $y$  is the literature density:

$$R^2 = \frac{\sum_{i \in x} (f(i) - \bar{y})^2}{\left( \sum_{i \in x} (f(i) - \bar{y})^2 + \sum_{i \in x} (y(i) - f(i))^2 \right)} \quad \text{Equation 3-3}$$

The low values of  $R^2$  especially for GAD67 indicates that the great variability of  $(x, y)$  could not be fully accounted for (roughly 25% of the variability accounted for GAD67, see Table 3-2).



**Figure 3-4 Linear fitting of marker intensity to cell density in cerebellum, isocortex, and the rest of the brain.**

Scatter plots of the PV+, SST+, VIP+ and GAD+ densities reported in literature (y-axis) according to the mean expression (x-axis) grouped by the main regions of the brain, respectively cerebellum, isocortex, and the rest of the brain. Each point represents an AV subregion and is color-coded according to the distinct levels of confidence (ratio of standard deviation over mean value) from literature data. The linear fit is represented with a black diagonal line. Each panel shows 95% of the available points. Regions that are purely inhibitory (i.e., based solely on the neuron counts obtained at step 2 of the pipeline), with no inhibitory neurons and regions with null region mean intensity are not displayed and were not considered for the fitting.



	Cerebellum		Isocortex		Rest	
	alpha	R <sup>2</sup>	alpha	R <sup>2</sup>	alpha	R <sup>2</sup>
PV	4.557×10 <sup>5</sup>	43.1%	2.964×10 <sup>5</sup>	46.4%	1.182×10 <sup>5</sup>	47.9%
SST	1.417×10 <sup>5</sup>	33.0%	2.523×10 <sup>5</sup>	50.6%	3.158×10 <sup>5</sup>	34.3%
VIP	7.903×10 <sup>4</sup>	26.4%	2.311×10 <sup>5</sup>	37.4%	1.154×10 <sup>4</sup>	56.5%
GAD	5.783×10 <sup>5</sup>	27.6%	3.458×10 <sup>5</sup>	37.0%	5.265×10 <sup>5</sup>	13.5%

Table 3-2 **Result of the linear fitting.**

The slope factor, alpha, of the linear fitting, and its coefficient of determination R<sup>2</sup> for each major region of the brain and for each genetic marker.

Thus, we use the fitted relation only to predict densities of GAD67+, PV+, SST+ and VIP+ in regions where no literature data are available. The standard deviations of these estimates are derived from the standard deviation of the alpha value of the fit (see Table 3-2).

For regions covered by the literature we instead use the average of all means and standard deviations of the reported values (see previous section). We assume indeed that the different literature values scatter around a common average.

Therefore, at this stage we have obtained unconstrained estimates of density (orig) and their standard deviation (σ) for PV+, SST+, VIP+ and GAD67+ neurons for all regions, i.e.  $\forall r \in \mathbf{R}$ :

$$\begin{aligned} orig_r &= \{nP_{V_r}; nSST_r; nVIP_r; nGAD_r\} & \text{Equation 3-4} \\ \sigma_r &= \{std(nP_{V_r}); std(nSST_r); std(nVIP_r); std(nGAD_r)\} & \text{Equation 3-5} \end{aligned}$$

orig and σ are lists of 3444 values: one per neuron type (PV, SST, VIP, GAD67) and per region of the brain (861 in AV1a).

### 3.2.6 Combination of neuron type densities

Since our unconstrained estimates of inhibitory neuron (orig and σ) densities are derived from the transfer functions or independent literature sources, their combination can lead to incongruent results. For instance, the unconstrained estimates of PV+ neurons might be greater than the estimates of GAD67+ neurons, which contradicts our assumption 4 (see Section 3.2.3). We therefore want to ensure that orig and σ match our assumptions from Section 3.2.3. To this end, we deduce a list of linear constraints based on Equation 3-1 and Equation 3-2,  $\forall r \in \mathbf{R}$ :

$$0 \leq nP_{V_r} \leq nNeu_r \quad \text{Equation 3-6}$$

And similarly, for nSST<sub>r</sub>, nVIP<sub>r</sub> and nGAD<sub>r</sub>.

$$-nNeu_r \leq nP_{V_r} + nSST_r + nVIP_r - nGAD_r \leq 0 \quad \text{Equation 3-7}$$

We also ensure that the consistency of the region hierarchy is maintained for every parent region ( $R_m \in \mathbf{R}$ ). The number of inhibitory neurons (and their subtypes) in  $R_m$  is equal to the sum of the corresponding estimates of  $R_m$ 's direct child regions in the region hierarchy ( $children_{R_m}$ ) plus the estimates in voxels of the AV labeled as belonging to  $R_m$  but none of its children ( $R_m \setminus child$ ):

$$nP_{V_{R_m \setminus child}} + \sum_{r \in children_{R_m}} nP_{V_r} = nP_{V_{R_m}} \quad \text{Equation 3-8}$$

and similarly, for nSST<sub>R<sub>m</sub></sub>, nVIP<sub>R<sub>m</sub></sub> and nGAD<sub>R<sub>m</sub></sub>.

Thus, whenever the constraints Equation 3-6, Equation 3-7, Equation 3-8 is violated, we rescale our unconstrained estimates (η).

We minimize the number of corrections required to find a solution for the BBcAv2 model through optimization. The number of corrections is defined as the sum of the distances between the unconstrained (η) and corrected values counterparts (x) divided by the standard deviation of the unconstrained value (σ):

$$\min. \sum_{r \in R} \frac{|x_r - \eta_r|}{\sigma_r} \quad \text{Equation 3-9}$$

This nonlinear function is convex which guarantees a global minimum. We convert it to a linear problem, by introducing a slack variable  $z$  so that  $\forall r \in R$ :

$$-z_r \leq x_r - \eta_r \leq z_r \quad \text{Equation 3-10}$$

Hence, our problem becomes:

$$\min. \sum_{r \in R} \frac{z_r}{\sigma_r} \quad \text{Equation 3-11}$$

st. Equation 3-6, Equation 3-7, Equation 3-8, Equation 3-10

The solution  $x$  is a vector with 3444 values and matches a total of 8613 linear constraints. An initial solution can also be found for each region, correcting the unconstrained estimates to match our constraints starting from leaf regions in the region hierarchy to the top-level brain regions (see Sup. Figure 3 and Sup. Figure 4). We use the simplex algorithm from the SciPy python library [79].

We find that for 16% of the regions of the brain (shown in colors in Figure 3-5AB), their  $\eta$  estimates are incongruent with the rest of the brain. Their corrected values  $x$  are therefore not falling within the range  $\eta \pm \sigma$ . In some regions of the striatum (including caudoputamen), the number of GAD67 neurons from  $\eta$  overshoots the estimated total number of neurons. Our optimization makes these regions fully inhibitory. This is in line with literature findings on the mouse striatum, which describe it as almost fully inhibitory [80]. Among the rest of the regions where significant corrections are needed, almost one-third are subregions of the hindbrain, for which we collected very few literature values. Conversely, few inconsistent first estimates are found in subregions of the isocortex, for which more literature data are available. The remaining inconsistencies for  $\eta$  may have diverse sources:

- They can be explained by a poor estimate of the densities from the transfer function.
- Divergent estimates from the literature for the same region might violate Equation 3-7. These regions would appear in yellow on Figure 3-5AB).
- Our assumptions may not be accurate in some regions of the brain.
- The densities of neurons from the step 2 of the pipeline (see Figure 3-1), used in Equation 3-6 and Equation 3-7, diverge locally from the literature findings

After our corrections, we obtain consistent density estimates for PV+, SST+, VIP+, and GAD67+ neurons for each region of the AV.

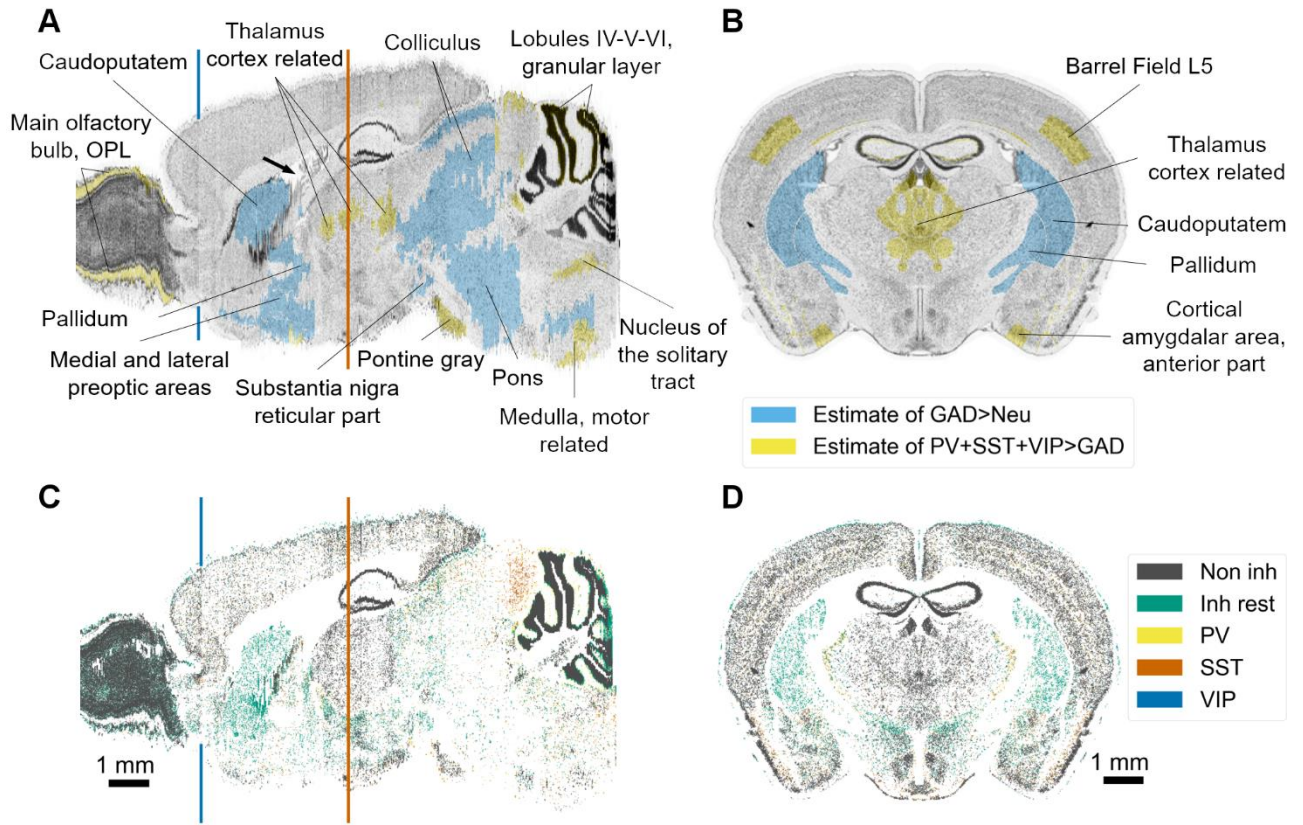


Figure 3-5 **Results of the density pipeline.**

(A) Sagittal and (B) coronal view of the Nissl reference atlas used for BBCAv2, showing cells of the mouse brain. Regions with high cell density appear in dark gray. Regions that required significant corrections (see Section 3.2.6) for their neuron subtype density estimates are colorized (Neu = neuron). The coronal slice chosen is displayed as an orange line on the sagittal slices. Arrow in black shows the location of the lateral ventricle of the mouse brain. The blue line behind the sagittal slice highlights the drop of Nissl expression coming from the original Nissl experiment from Dong [7]. (C). Sagittal and (D). coronal view of the BBCAv2, showing the positions of the different types of neurons. PV+, SST+ and VIP+ cells appear respectively in yellow, dark orange and blue. The rest of the GABAergic population is color-coded in green and the remaining neurons in gray. The variation of the distribution of neurons is following the original distribution of Nissl expression. The Nissl expression drops significantly after the blue line which leads to a similar decrease of cell counts in the resulting BBCAv2.

### 3.2.7 Placement of the cells within the mouse brain volume

At step 4 of the pipeline (see Figure 3-1), we convert our density estimates into a list of cell positions for each region, using an acceptance-rejection algorithm. For each region, the algorithm samples voxels, according to the cell density distribution (computed at step 2), and assigns cells to them, until it reaches the target total number of cells in that region. The same algorithm can then be applied to label cells as neurons or glial cells as described in Erö et al. [8]. We further subdivide neurons into PV+, SST+, VIP+, and GAD67+ neurons, following our density estimates. In brief,  $\forall r \in \mathbf{R}$ , we select uniformly neurons within the voxels of region  $r$ , that we label as inhibitory, until the region contains the target number of  $n\text{GAD}r$  inhibitory neurons. Similarly, we distribute the  $n\text{PV}r$  neurons, in voxels of  $r$  containing GAD67 neurons,  $n\text{SST}r$  in voxels of  $r$  containing non-PV inhibitory neurons, and  $n\text{VIP}r$  in voxels of  $r$  containing the remaining inhibitory neurons. The non-inhibitory neurons are labeled as excitatory or purely modulatory neurons following Equation 3-1.

Finally, we assign to each cell a uniformly random 3D position within the boundaries of its voxel. The result of this procedure is shown in Figure 3-5CD.

### 3.2.8 Assign me-types to neurons in the whole brain model

At this stage of the pipeline, each of our neuron has a molecular identity (excitatory, inhibitory, pv+, sst+, vip+). We want now to refine these types and integrate literature knowledge on the morphological and electrical properties of these neurons. In the isocortex, we used to override the results of the BBCAv1 with the distribution from Markram et al.'s circuit [4]. This was changing the ratio of inhibitory and excitatory neurons that we found and uniformizing it to every region of the isocortex. Instead, we want to leverage the results of the BBCAv2 to predict the distribution of morphological (m-type) and electrical (e-type) neuron types in the Isocortex. To do so, we need to map our molecular types (excitatory, inhibitory, pv+, sst+, vip+) to the me-types provided by Markram et al.

We used Roussel's pipeline to refine our labeling of the GABAergic neurons in the isocortex [63]. This pipeline uses datasets from Gouwens et al. [81] (AIBS datasets) and Markram et al. [4] (BBP datasets) and creates a mapping between molecular cell identities and morpho-electrical cell type features. In brief, the different steps of this pipeline are:

- Morphological and electrical (me-) features were extracted from the AIBS and BBP datasets (Figure 3-6AB). These me-features were normalized to compensate for the variability due to animal species, cortical subregions differences, or neuron positions within layers.
- Normalized me-features were integrated into a knowledge graph and features common to the two datasets were selected (Figure 3-6CD).
- Common me-features were weighted and clustered to recover cell types categories using Ward hierarchical clustering [82] (Figure 3-6E).
- Cell types were regrouped by optimization to maximize the precision of molecular identities predictions (Figure 3-6F).
- The final cell clusters were used to create a probability map between the BBP me-models and a molecular identity.

To each cortical neuron we assigned an me type label based on the mapping from Roussel.

In the cerebellar cortex, we also refined the neuronal composition using the tools we defined previously. The Purkinje layer defined in Section 2.2.4 contains solely Purkinje cells that are arranged in arrays in the parasagittal plane [14]. Based on the cell density numbers reported by Lange [83] and Förster [84], the density within these voxels will be equal to approximately 21000 cells per mm<sup>3</sup>. We then assigned a m-type to each neuron of the granular and molecular layer, based on our new inhibitory excitatory neuron distribution. We further extended the cell types from Casali et al. [14] to include the Unipolar brushed cells and the Lugaro cells which are respectively excitatory and inhibitory neurons located in the granular layer [85], [86]. We computed the number of each neuron type based on the following recipe:

- The densities of Unipolar Brushed Cells in each region were extracted from Sekerkova et al [85]. The rest of excitatory neurons within the granular layer were labeled as Granule Cells.
- The number of Lugaro cells can be processed from the number of Purkinje cells [86]. The rest of the inhibitory neurons are Golgi cells.
- We split the Molecular layer in 2 using the orientation vector fields computed at Section 2.3.3. Following this axis from the cerebellar fiber tracts towards the exterior of the molecular layer, one third of the molecular layer contains Basket cells, the rest are Stellate cells [14].

Except for the Lugaro cells and the Unipolar brushed cells, the paper from Casali et al. provides also a set of parameters for integrate and fire models, which could be implemented in our model. However, this new neuronal composition refinement does not take into account the PV, SST and VIP subdivision that we obtained with the new pipeline but could be integrated when a mapping between these molecular types and the m-types defined previously can be found in literature. Similarly, in the basal ganglia complex, we used the circuit of Lindahl et al. [19] to obtain a more fine-grained distribution of the neuron types in these regions. The neuronal composition within the thalamus VPM, RT and POM regions was updated based on results of [48]. In the hippocampus CAs, we used the recipe from [5]. The latter two recipes altered the Excitatory/Inhibitory ratios in these regions as for the cerebellar cortex. In a future version of the cell atlas, these literature findings should be implemented a priori to prevent the overriding of the results found earlier. For the rest of the brain regions, whenever no reports in literature provided a complete description of the neuron types within, we kept the cell type labels from the BBCAv2.

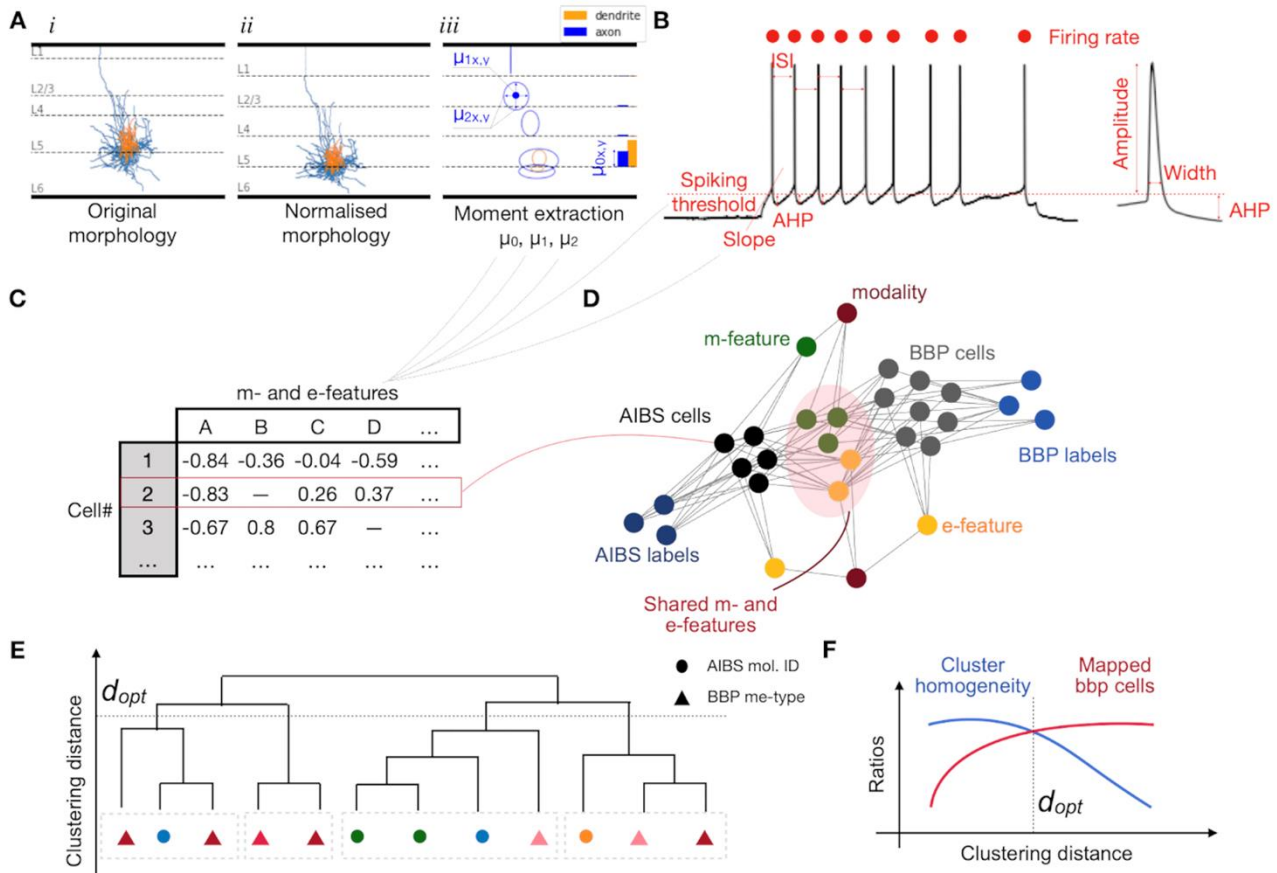


Figure 3-6 **Overview of Roussel et al. pipeline.**

(A) Morphological feature extraction. Original morphologies (i) were realigned to be contained within cortical limits. X and Y coordinates were normalized by cortical depth and corresponding layer thickness, respectively resulting in normalized morphologies (ii). Density moments of order 0, 1 and 2 were extracted separately for axons and dendrites and for each layer individually (iii). Moments of order 0 can be related to total neurite length (bar chart on the right side of the panel) whereas moments of order 1 and 2 can be linked to mean and standard deviation of the density (ellipses). Axons in blue, dendrites in orange.

(B) Electrophysiological recordings from step currents were used to extract a collection of electrophysiological features.

(C) All extracted features were stored in tables and used to produce a knowledge graph (D) grouping all information that can be extracted from both datasets. Each cell is represented by a root node, that is linked to nodes representing the different properties of that cell (morpho-electrical features or other information).

(E) We used ward hierarchical clustering results to group common me-features and (F) we established the optimal clustering distance  $d_{opt}$  as a compromise between averaged cluster homogeneity with respect to molecular IDs (blue curve) and ratio of BBP cells belonging to mixed clusters (red curve).

### 3.3 Results

#### 3.3.1 An improved computational method to build an atlas of inhibitory neurons in the mouse brain

In this study, we present a substantially improved workflow to produce the second version of the Blue Brain Mouse Cell Atlas workflow (BBCAv2). The new workflow replaces the manual alignment of ISH datasets with the image alignment algorithm developed by Krepl et al. [36]. The pipeline also includes a new method to estimate cell densities from a large body of literature data. In total, we integrated values from 56 different literature sources.

All steps of this process are guided by a set of assumptions about the cellular composition of the mouse brain. Using our improved workflow, we are not only able to re-estimate the densities of inhibitory neurons in the mouse brain, but also to estimate the densities of their various subtypes. The result is an updated mouse cell atlas which provides density estimates and their standard deviations for each cell class in all regions of the CCFbbp AV. We additionally derived from our cell atlas densities a 3D position and a type label for each cell within the mouse brain (see Figure 3-5CD). The same methods can also be applied to other cell types to deepen our knowledge on cellular composition of the mouse brain.

#### 3.3.2 Distribution of inhibitory neuron subtypes for the whole mouse brain

The new cell atlas has updated estimates for the densities and numbers of inhibitory neurons, based on more literature data. In addition, the new atlas differentiates PV+, SST+ and VIP+ neurons and provides their densities and absolute numbers for each region of the mouse brain. An overview of the results is shown in Figure 3-7 and Table 3-3. We estimate that the mouse brain holds approximately 14.55 millions of inhibitory neurons in total (see Table 3-3). GABAergic neurons make up 20.27% of the total neuron population, which is significantly larger than the previous estimate of 15.68%, using the BBCAv1 workflow. We find that 3.57% of neurons are PV positive, 3.20% are SST positive, and 0.63% are VIP positive (see Figure 3-7A). This leaves a residual GABAergic neuron population of 12.87%, which represents almost two-thirds of all inhibitory neurons.

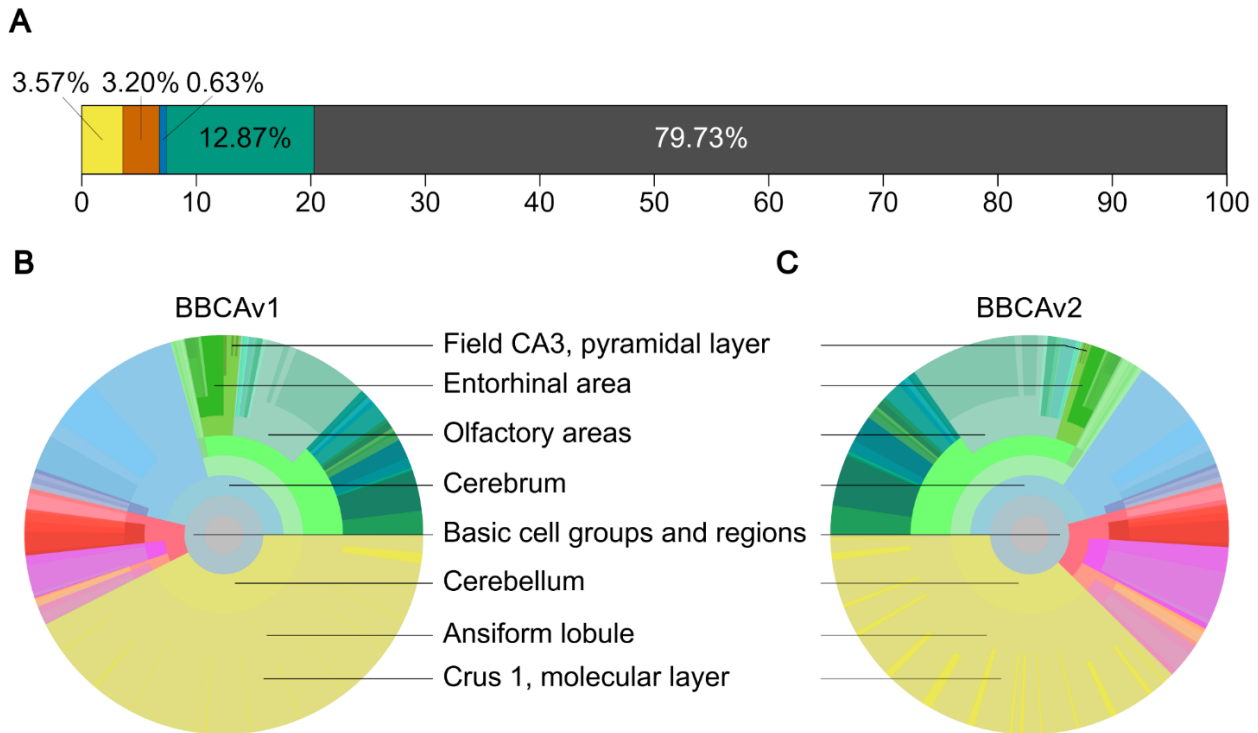


Figure 3-7 **Distribution of the GABAergic neurons in the mouse brain.**

(A) Ratios of PV+, SST+, VIP+, GAD67+, and remaining neurons of the mouse brain. InhR cells appear in green, PV+ in yellow, SST+ in orange, and VIP+ in blue. The remaining neurons appear in gray.

(B), (C) Circular distribution of GAD67+ neurons in different regions of the mouse brain according to different parcellation schemes ranging from coarse at the center to fine-grained at the periphery (colors and disposition similar to Figure 3-3A). (B) displays results of BBCAv1 and (C) results of BBCAv2.

Compared to BBCAv1 (Figure 3-7B), the number of inhibitory neurons in BBCAv2 (Figure 3-7C) is larger in almost all regions of the brain (except notably, the hippocampal formation and the striatum). The estimated number of inhibitory neurons in the cerebellar cortex granular layer increases from 203,000 in the previous pipeline to 867,000. Similarly, in the isocortex, this number rises roughly by 53% (see Table 3-3). Finally, the BBCAv2 estimates 2.05 million inhibitory neurons in the olfactory areas.

Brain Region	BBCAv1	BBCAv2					
	Inh.	Inh.	$\frac{\text{Inh}}{\text{Neu}}$ (%)	$\frac{\text{PV}}{\text{Inh}}$ (%)	$\frac{\text{SST}}{\text{Inh}}$ (%)	$\frac{\text{VIP}}{\text{Inh}}$ (%)	$\frac{\text{Rest}}{\text{Inh}}$ (%)
Isocortex	1.44	2.20	24.7	23.2	25.0	12.6	39.2
Olfactory areas	1.45	2.05	25.1	1.3	13.8	4.7	80.2
Hippocampal form	0.51	0.49	9.9	12.6	33.9	9.0	44.5
Cortical subplate	0.12	0.22	43.8	2.6	28.3	3.7	65.4
Striatum	1.72	1.40	81.4	1.2	17.5	0.2	81.1
Pallidum	0.18	0.23	91.8	6.9	28.7	0.2	64.2
Thalamus	0.19	0.26	18.9	15.2	32.7	0.0	52.1
Hypothalamus	0.42	0.50	41.4	3.0	29.0	0.4	67.6
Midbrain	0.39	0.99	76.4	12.6	37.4	0.9	49.1
Hindbrain	0.26	0.66	62.5	23.9	38.2	0.6	37.3
Cerebellum	4.79	5.47	13.0	28.8	0.8	0.1	70.3
Whole Brain	11.25	14.55	20.3	17.6	15.8	3.1	63.5

Table 3-3 Estimates of inhibitory neurons in the mouse brain in millions and percentage (%) subtypes.

Inhibitory neuron counts (in millions) extracted from the cell atlas generated using the BBCAv1 and BBCAv2 pipelines. The second column of the BBCAv2 corresponds to the ratio of inhibitory neurons according to the total number of neurons in this region. The following columns display the proportion of each inhibitory neuron subtype according to the inhibitory neuron population.

### 3.3.3 New Excitatory/Inhibitory ratios for the isocortex

The BBCAv2 also provides new estimates for the ratio of inhibitory/excitatory neurons (see Figure 3-8). For example, we predict that 25% of the neurons of the mouse's isocortex to be inhibitory (see Table 3-3 and Figure 3-8B). Thus, we estimate a higher proportion of inhibitory neurons than previously reported. For rodents, literature estimates are around 20% inhibitory neurons [4], [87].

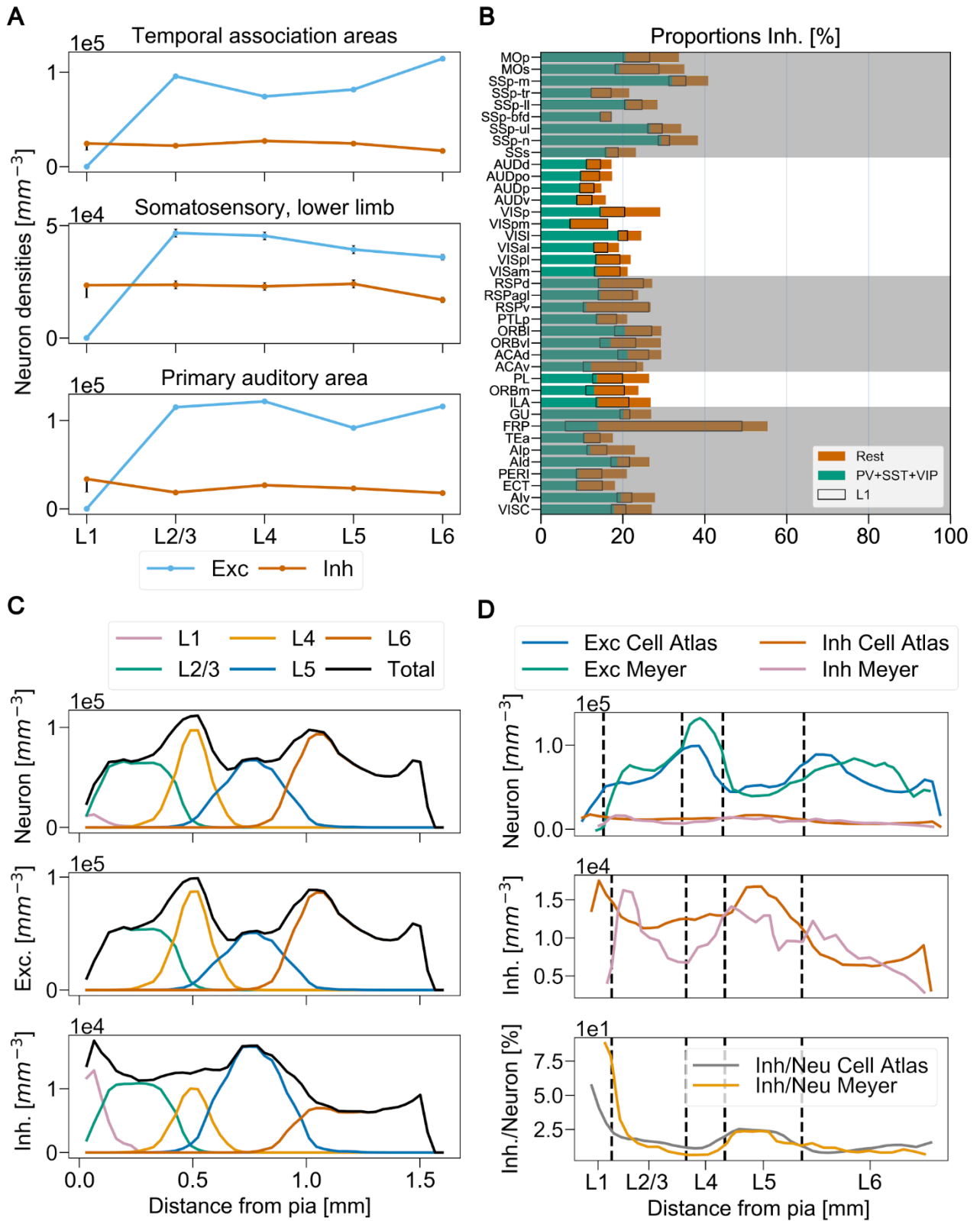




Figure 3-8 **Cortical excitatory/inhibitory ratios.**

(A) Relative cortical layer cell densities from the primary somatosensory cortex lower limb (SSp-ll), primary auditory cortex (AUDp) and temporal association areas (TEa). Blue lines correspond to excitatory neuron densities (Exc) and the orange lines inhibitory neuron densities (Inh) in cells/mm<sup>3</sup>. Final confidence intervals are also displayed for each layer as vertical lines.

(B) Ratios of inhibitory/excitatory neurons (Inh/Exc) across anatomical regions of the isocortex, expressed as percentage and arranged in five subnetworks based on Kim et al. [16]. Acronyms correspond to the AV naming convention [34] (see Sup. Table 1). The portion of inhibitory neurons that belongs to layer 1 is shown with a black contour.

(C) Somatosensory barrel field neuron distributions according to pia. Each layer is represented by a distinct color and the total density according to distance to pia appears in black.

(D) Somatosensory barrel field excitatory/inhibitory neuron distributions (Exc/Inh) according to pia compared to the results extracted from Figure 2-5EFG from Meyer et al. [87]. AV mean layer limits are represented as dash lines and aligned to the ones provided by Meyer et al.

The poor alignment of the AV to the Nissl dataset counterpart (including the layer 1 layer 2 boundary, see Section 2.2.3) may explain this result. It is artificially raising the density of neurons in layer 1, which in turn, increases the ratio of inhibitory neurons in the isocortex (black contour in Figure 3-8B). Additionally, the densities of PV+, SST+ and VIP+ neurons, mostly constrained by Kim et al. [16], are providing a lower limit for the total number of inhibitory neurons in each region (green bars in Figure 3-8B). To improve on the ratios of inhibitory neurons in the isocortex we should therefore check its densities of neurons and our predictions of GAD67 neurons in those regions. We also notice that the estimated densities of inhibitory neurons are quite similar from layers 2 to 5, and across isocortex subregions (Figure 3-8A and Sup. Figure 5). As an additional validation, we check if our model can reproduce the findings of Meyer et al. on rats in the barrel cortex [59]. To do so, we computed the neuron density profiles according to the distance between their soma position and the pia (see additional methods and Figure 3-8C). The direct comparison of our density profile and the results from Meyer et al. are shown in Figure 3-8D [87]. The distributions of the excitatory and inhibitory neuron populations according to their distance to the pia are similar despite the cross-species comparison.

### 3.3.4 New distribution of me-types for the isocortex

Using Roussel et al. pipeline [63], we were able to obtain a new distribution of neuron me-types within the isocortex. These results can therefore be directly compared with the ones obtained with Erö's pipeline (see Section 3.1). Sup. Figure 9 and Figure 3-9 show the distribution of the neuron me-types in the somatosensory cortex lower limb in respectively BBCAv1 and BBCAv2.

First, as we discussed in Section 3.3.3, the inhibitory neurons proportion has increased when compared to BBCAv1 (see Sup. Figure 9AB and Figure 3-9AB). We notice that this change is more important in the subgranular layers (layer 4 to 6). As we discussed in Section 3.3.3, the density of inhibitory neurons seems to be similar from layer 2 to layer 6, but the densities in the subgranular layers is lower than in L2/3. This can partially explain this change in inhibitory proportion. Second, we observe that the populations of Martinotti cells (MC) and the nested and large basket cells (NBC and LBC) seem to take a larger portion of the m-type distribution when compared to BBCAv1 (see Sup. Figure 9C and Figure 3-9C). As discussed in Roussel et al. [63], this might be linked to an over-representation of these neurons in the original BBP datasets used to construct the cell type to molecular identity mapping. Similarly, an under-representation of the Lamp5 population in the original AIBS datasets. Third, the e-type distribution changed drastically between BBCAv1 and BBCAv2. We observe that the inhibitory e-type population is more evenly distributed (see Sup. Figure 9D and Figure 3-9D). These differences in e-type distribution cannot be fully explained by the m-type changes previously observed. Indeed, the Markram et al. [4] distribution of m-to-e-type distribution would not explain the more important proportion of the bursting Stuttering (bStut) population solely based on an increase of MC, SBC, and LBC neurons. This indicates a more profound change in the proportion of m-to-e-type in SSCtx-LL in our model. This can be due to the label regrouping of the inhibitory e-type labels from the BBP datasets under the Fast Spiking (FS), non-Fast Spiking (nFS), irregular spiking (IR) and adapting (Adapt) categories in the AIBS dataset. This regrouping mapped most of the continuous non-Accommodating (cNAC), delayed non-Accommodating (dNAC), continuous Stuttering (cSTUT) and continuous Stuttering (dSTUT) types under the FS category. All these e-type populations were reduced with the new mapping, except for the cSTUT population, which might indicate that the proportion of the FS category was overall reduced in the new model compared to the old distribution.

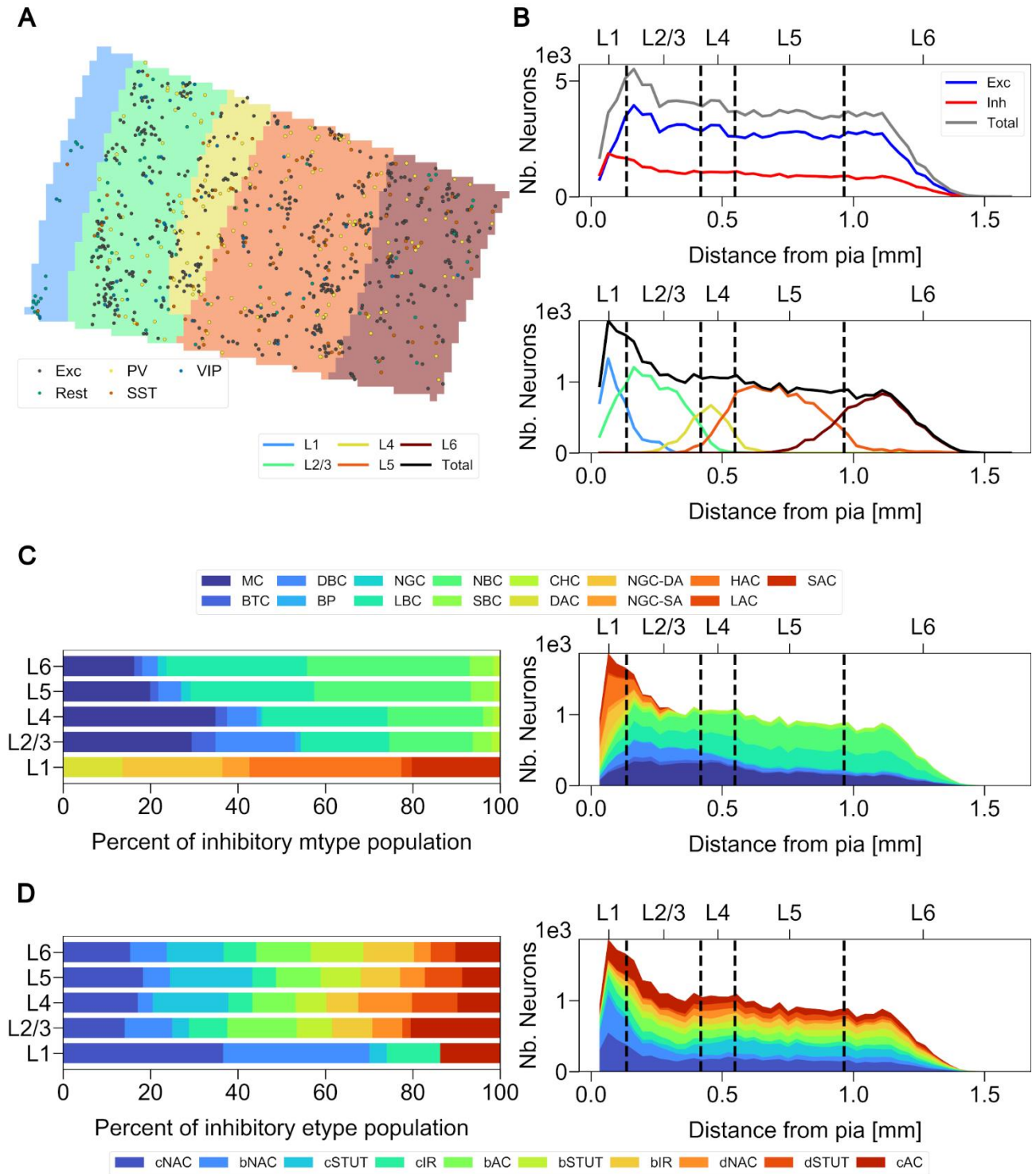


Figure 3-9 **Distribution of neurons in the BBcAv2 Somatosensory cortex lower limb.**

A. Coronal view of the BBcAv1 neuron distribution in the Somatosensory cortex lower limb region (SSCtX-LL). This panel shows the positions of distinct types of neurons in the different layers. Inhibitory and excitatory cells appear respectively in red and blue.

B. Distribution of neuron counts in SSCtX-LL according to their distance to the pia. The excitatory distribution is displayed with the blue line, the inhibitory population with the red line and the sum of the two is shown in gray. AV mean layer limits are represented as vertical dash lines.

C. Distribution of Inhibitory Morphological types (m-types) in SSCtX-LL according to their layer (left panel) and their distance from the pia (right panel). Each m-type population is represented with the same color in both panels. The left panel shows the proportion of the inhibitory m-types in each layer as percentage. The right panel shows the cumulative distribution of the inhibitory m-types according to their distance to the pia. AV mean layer limits are represented as vertical dash lines.

D. Distribution of Inhibitory Electrical types (e-types) in SSCtX-LL according to their layer (left panel) and their distance from the pia (right panel). The panels are analog to C.

### 3.3.5 Validation of the pipeline

In our revised workflow for BBCAv2, we used a linear transfer function to map ISH expression levels to literature cell density to then estimate the cell densities in regions where no literature value is available. The fidelity of the transfer function depends on both the quality of the ISH images and the quality of the underlying literature estimates of the cell densities. As we highlighted in Section 3.2.2, the genetic marker datasets are prone to artifacts due to slicing. Also, their coverage of the mouse brain is not perfect (see Figure 3-2D). Combining multiple datasets of the same genetic marker can blur the spatial details of brain structures. For each marker, the AIBS website hosts similar ISH experiments in sagittal (Figure 3-11A) and coronal (Figure 3-2A) slices. We compare the results of our pipeline, using two different coronal ISH datasets for GAD67 and find very similar results in terms of fitting, distribution of inhibitory neurons and remaining errors after corrections (see Sup. Table 2). To blur the individual artifacts of the ISH datasets, we combine these datasets after realignment, taking the mean of their expressions. We then linearly fit the marker expression from the resulting dataset to literature densities, but we find no significant improvement. This result can be explained by the slicing of these two experiments which, in both cases, is not fully covering the tips of the olfactory bulbs and the cerebellum as illustrated in Figure 3-2D.

Next, we explore the impact of the amount of literature values available on the quality of the transfer function. We do this experimentally by running our workflow with a fraction  $f$  of the available literature points (70% up to 95%). For each fraction  $f$ , we perform 20 separate experiments, each with a different random subset of literature values. Thus, for each fraction  $f$ , we obtain 20 different estimates of the cell density in each brain region  $r \in \mathbf{R}$ . We also compute the standard deviation of the resulting density values for each region ( $\sigma(r, f)$ ). As we increase the number of literature values that are contributing, we expect the standard deviation of the density estimates to decrease. We observe that the average of all standard deviations for each marker gets smaller as more literature values are integrated (see Figure 3-10A).

For each region  $r$ , we then use all  $\sigma(r, f)$  to extrapolate the standard deviation of the density for  $f=100\%$ . This standard deviation  $\sigma(r, 100\%)$  represents the confidence of our pipeline in our final predicted densities for each region.

Surprisingly, the quality of the fitting does not significantly improve with more literature values added to the pipeline (see Figure 3-10C). This result indicates that the coefficient of determination is mostly influenced by the neuron distribution obtained at step 2 (see Figure 3-1), the ISH experiments or the method to group the regions (here cerebellum, isocortex and rest, see Figure 3-4).

We also test the capacity of the BBCAv2 to predict values of literature. We select 90% of the literature points available, and then compare the generated densities to the remaining points (see Figure 3-10B). Most of the density values produced with this method fall within the confidence interval of their literature counterparts. This workflow is therefore capable of predicting densities of cells within range of what literature provides.

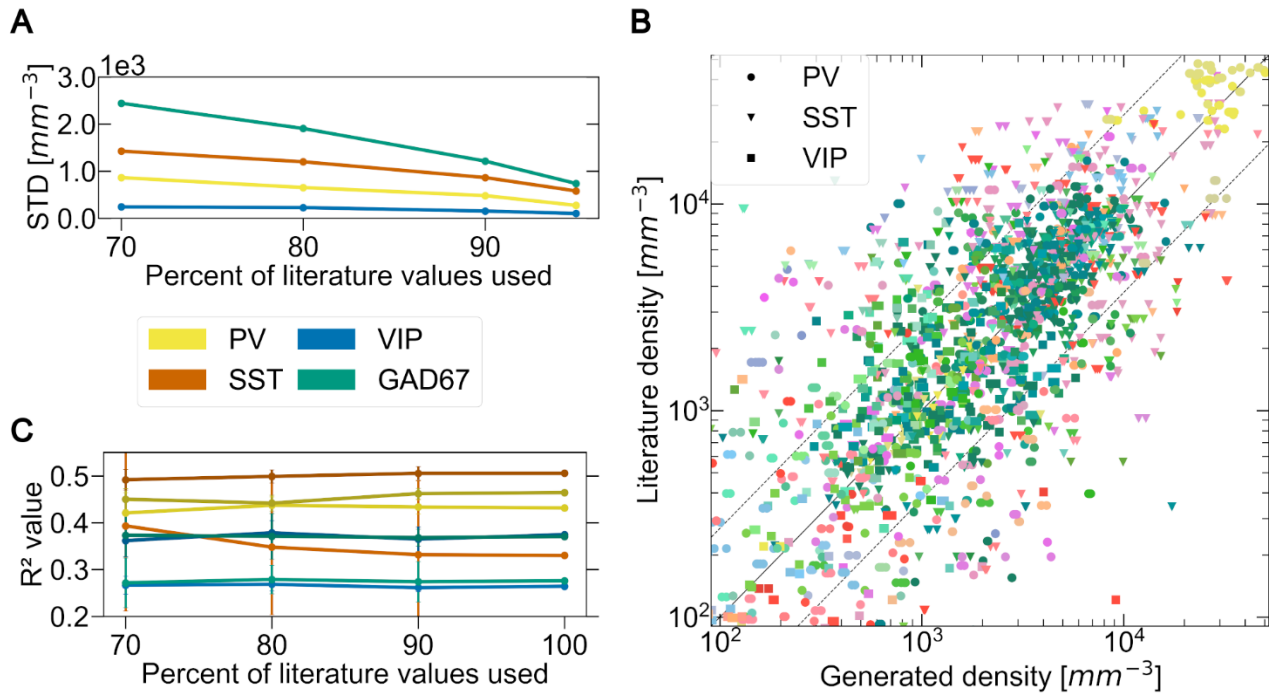


Figure 3-10 **Impact of the amount of literature on the BBCAv2.**

(A) Evolution of the mean standard deviation of each genetic marker (shown in distinct colors) using different percentages of the total amount of literature values available.

(B) Comparison of the generated density values of the BBCAv2 with their literature counterparts using 90% of the total amount of literature values available (20 trials). The color encodes the brain regions according to the AV, while the shapes of the points encode for cell types. The middle line delimits equal quantities, while the dashed line shows the average deviation of 2.7-fold between literature values reporting on the same region for PV, SST and VIP neurons.

(C) Evolution of the coefficient of determination  $R^2$ . Each line corresponds to the  $R^2$  value of the fitting performed using different percentages of the total amount of literature values available. The color of these lines is linked to the gene chosen and their color shade corresponds to their fitting group (here isocortex and Rest, as in Figure 3-4).

As an additional qualitative validation, we compare our generated GABAergic neurons positions with a sagittal slice of GAD67 expression from the AIBS (see Figure 3-11AB) and find a similar whole brain distribution. Some of the visual differences between Figure 3-11A and Figure 3-11B can be explained by the fact that in the image from the BBCAv2 (Figure 3-11B), the soma sizes are uniform across all cells, while in the brain, soma sizes differ considerably between cell types and brain regions (Figure 3-11A). Moreover, the AV1a is based on the coronal Nissl experiment from the AIBS (see Section 2.1) which is not well covering the tips of the cerebellum and the olfactory bulbs (see Figure 3-5A) and deforming them. However, this comparison highlights the fact that in our model the thalamus and the cerebellum have too many inhibitory neurons. Additionally, the granule layer of the dentate gyrus is clearly visible on the original experiment but not in the BBCAv2 brain slice (center black arrow). Our density estimates for this region ( $5403 \text{ mm}^{-3}$ ) are, however, in the range of the literature [43], [88]. The granule layer of the dentate gyrus is very dense [8], and Zeisel et al. [66] have shown that its excitatory neuron population expresses GAD67 at a low level, but not GAD65. A closer inspection of the GAD67 and GAD65 ISH brain slice images allowed us to identify two types of cells expressing the markers. The cells with the largest radius and the strongest level of expression of GAD67 and GAD65 are sparse and their description corresponds to the inhibitory neurons, as in Jinno et al. study [43]. The other population of cells are smaller in diameter and their expression of GAD67 is lower compared to the other population. This population is additionally not visible in the GAD65 brain slice, which indicates that it corresponds to the excitatory neuron population.

In Figure 3-11, the original slice in panel A shows spatial detail with higher resolution than our reconstruction in panel B. For example, in the olfactory bulb, panel A shows the layering of cells, as we do in the cerebellum and hippocampus. The spatial resolution of our cell atlas is lower because the density is assumed uniform within regions at the lowest level of the AIBS region hierarchy (assumption 5), following the estimated cell density.



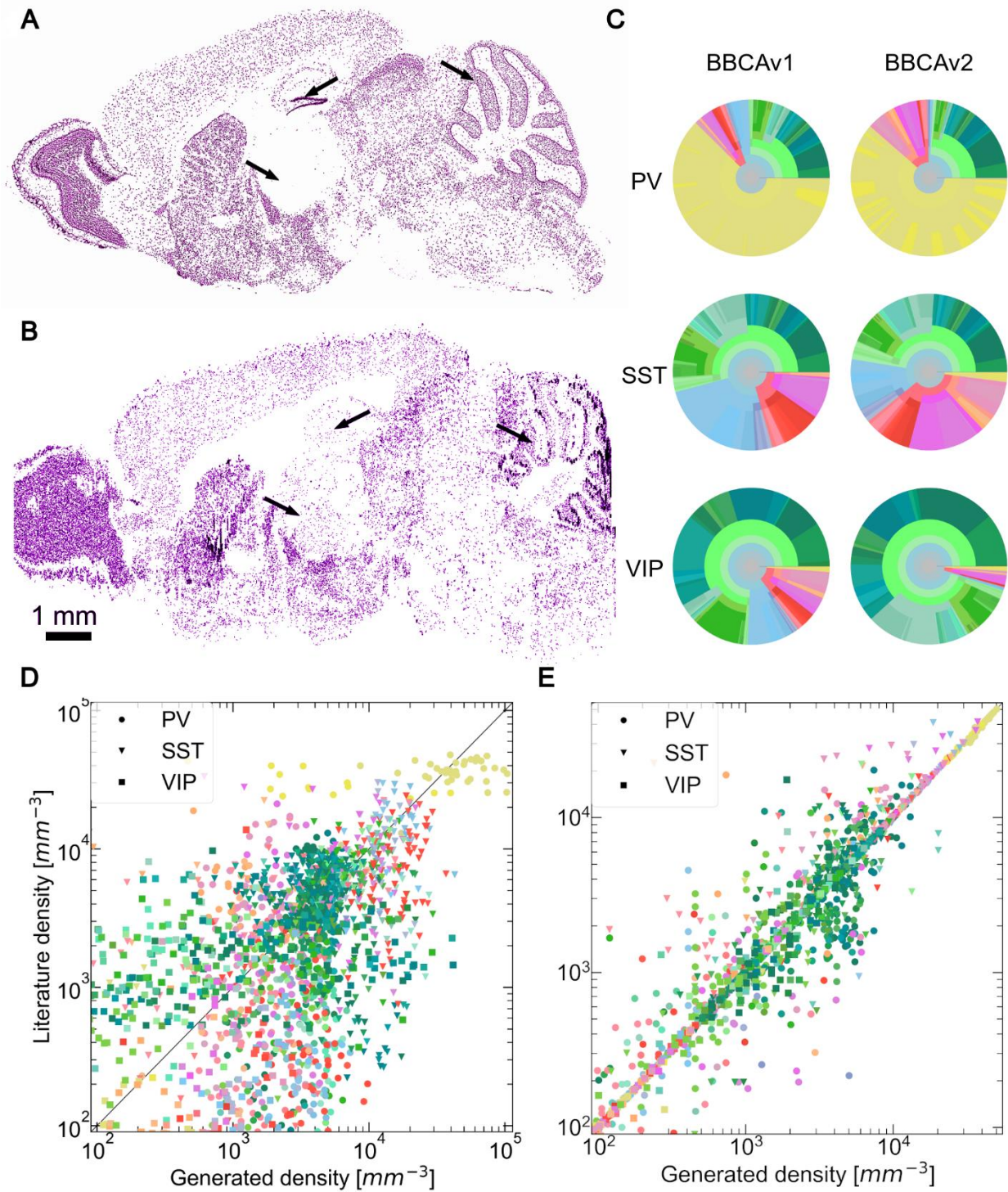


Figure 3-11 **Qualitative validation of the BBCAv2 and comparison with BBCAv1.**

GAD67 sagittal brain slices comparison between (A) the filtered image from an ISH experiment by the AIBS at the top and (B) the BBCAv2 predicted GABAergic neurons positions at the bottom. Cells reacting to the marker are displayed in purple. Blank regions in our model correspond to fiber tracts regions which host no neuron. This slice does not take regional soma sizes into account. Most of the density features of the real experiment can be found in the BBCAv2 slices. Black arrows point to the thalamus (left arrow), molecular layer of the dentate gyrus (center arrow) and granular layer of the cerebellum (right arrow) of each brain.

(C) Circular distributions of the densities of PV+, SST+ and VIP+ neurons computed according to the Erö et al. method on the left [8] and our method on the right (colors as in Figure 3-3A).

Distribution of PV+, SST+ and VIP+ neuron density values reported in literature against generated densities, using the BBCAv1 method (D) and the BBCAv2 method (E), for similar regions. The color encodes the brain regions according to the AV, while the shapes of the points encode for cell types. The middle line delimits equal quantities.

We also compare the estimates of PV, SST and VIP densities from the BBCAv2 against similar estimates with the BBCAv1 [8]. For the BBCAv1, we use the global number of PV, SST and VIP neurons in the mouse brain from Kim et al. [16] and the PV, SST, and VIP realigned datasets (see Section 3.2.2, Figure 3-11C). The BBCAv1 maintains the global number of inhibitory subtypes expected in the whole brain at the cost of losing the local distribution (see Figure 3-11D). For instance, Kim et al. [8] found that the cerebral nuclei regions contain 62 VIP neurons per mm<sup>3</sup>, while the BBCAv1 predicts it to contain 963 VIP neurons per mm<sup>3</sup> (see Figure 3-11C, cerebral nuclei regions appear in light blue). The BBCAv2 method, on the other hand, aims to maintain local estimates from literature in a coherent atlas (see Figure 3-11E). It is therefore usually underestimating the densities of each neuron type according to literature since the neuron and inhibitory neuron densities act as maximum constraints. It is therefore usually providing density estimates closer to their literature counterparts for each neuron type.

Our model integrates literature values in the coherent framework of the cell atlas, with constraints and assumptions that we control and that we can refine whenever more data become available about the mouse brain.

## 3.4 Discussion

We have described a novel approach to estimate cell densities in the mouse brain based on literature findings (see Section 3.2.4), ISH experiment data from the AIBS (see Section 3.2.2) and clear assumptions (see Section 3.2.3). This allows us to provide new estimates of inhibitory neuron counts for each region of the mouse brain, including PV+, SST+, VIP+ neurons, and remaining inhibitory neuron subtypes, in a consistent framework: The Blue Brain Mouse Cell Atlas (BBCAv2). BBCAv2 extends the results of BBCAv1 and deepens our knowledge of the cellular composition of the mouse brain. Additionally, the BBCAv2 can be easily adapted to estimate the densities of other cell types, providing more details about the composition of the brain in terms of morphological, molecular, or electrical properties. Further analyses can be carried out to study ratios and correlations of the different cell types of the brain.

### 3.4.1 A new approach to integrate literature findings into the Blue Brain Cell Atlas

To compensate for the great variability of cell density measurements in literature and to avoid a bias towards any particular finding, we decided to use as many data sources as possible (see Section 3.2.4) in our workflow. The BBCAv2 therefore becomes a reflection of our current knowledge about cell composition in the mouse brain. However, this makes it also more difficult to validate, since only limited data or studies remain available to compare our results to. Moreover, any additional literature value that we could use to validate our counts might itself be as trustworthy as the literature values used to build our model. On the other hand, qualitative validations can highlight inconsistencies in our results as shown in Figure 3-11AB.

As we will discuss, obtaining ground truth for densities of cells in the mouse brain is an exceedingly challenging task. Density estimation is indeed subject to many diverse sources of variability such as data acquisition techniques, brain region delineation and inter-subject variability as well as a lack of knowledge about cell types in several regions of the brain (see Figure 3-3A and Table 3-1). We need to be aware of all these biases and potential errors that may influence our estimates. In this paper, we therefore tried to clearly state all our assumptions, to measure the impact of each dataset that we used, and to list all the remaining problems that still need to be solved.

### 3.4.2 Effect of alignment impact on cell density estimation

Both Blue Brain Mouse Cell Atlas pipelines are based on a pair of reference atlas datasets from the AIBS which exist in multiple versions. Prior to refining the BBCAv1, we had to measure the reliability of each reference atlas version available. We assessed newer versions of the AV and analyzed their impact on cell densities calculation. We found that the best cell density estimates can be obtained if the pair of reference datasets (Nissl volume and AV) is perfectly aligned to each other. For example, if the border between two regions with significantly differing densities (e.g., isocortex layer 1 and layer 2, see Figure 2-5AC) is not perfectly aligned with the corresponding Nissl, then the densities in the particular regions will be over- or underestimated. Hence, the CCFv3 AV, while depicting a smooth average brain volume, provides the worst density results when combined with Nissl2.

The Nissl reference volume from Dong [7] shows the cell distribution of a single individual and the shape of its brain has been stretched and sometimes torn by the slicing, which yields a very noisy dataset. The derived annotation atlas had to follow this roughness with sometimes discontinuous region borders. The resulting AV are therefore less attractive for circuit modeling. We can still benefit from the best of each dataset, computing the densities using the CCFv2 or CCFbbp and then populate the CCFv3 smooth brain volume with these estimates. Cell distribution in the lowest level of the region hierarchy will be considered homogenous. Potential inner region gradient can be kept if a common vector base is defined in each AV version.

### 3.4.3 Challenges linked to ISH marker expression datasets

Another crucial stack of input data consumed by the BBCAv2 pipeline includes the filtered ISH images of the mouse brain from the AIBS. These datasets were realigned at step 1 of our method and used at step 3 to obtain density estimates in the regions missing literature values.

First, we considered the ISH image slices to be perfectly vertical to simplify their realignment to Nissl and the interpolation between them. This assumption is however not true and the AIBS provides the position of the image corners in each ISH experiment metadata. In a future iteration of our pipeline, we will consider integrating each ISH slice at its exact position as it might help the algorithm to find the right landmarks to realign the images to the Nissl counterpart.

The ISH experiments cover partially the whole mouse brain, with one slice stained per 200  $\mu\text{m}$  in the rostro-caudal axis (see Figure 3-2D) [21]. Additionally, they represent different mouse individuals whose neuron distribution might vary a lot with respect to one another. As we discussed in Section 3.3.5, the combination of similar ISH datasets can improve their coverage of the brain and blur out the remaining artifacts. Furthermore, ISH experiments performed with a sagittal slicing covers regions not included in the coronal experiments and their addition will improve the ISH whole brain coverage. Using these datasets requires the realignment of the sagittal images of the inhibitory markers to the Nissl volume, using the deep learning algorithm described from Krepl et al. [36]. The usage of these datasets is currently investigated for a future update of the pipeline.

The marker datasets, despite having been processed by the AIBS, are still subject to artifacts and noise which will artificially raise or lower our cell counts estimates (see Figure 3-2A). To counter these effects, Erö et al. [8] used the *Nrn1* marker, as it is expressed in excitatory neurons, to limit the expression of GAD67. The expression of GAD67 was here scaled according to the expected ratio between excitatory and inhibitory neurons from Kim et al. [16] and normalized according to the sum of the expression of GAD67 and *Nrn1*. The GAD65 marker has also been reported to be expressed in GABAergic cells [89]. Similarly, we can extend our approach, knowing the density of neurons expressing *Nrn1* and GAD65 in regions of interest, to correct our estimates of GABAergic neurons.

We also chose to use an all-or-none filter on the filtered images because we were interested in counting all the cells expressing the marker. However, we could deduce counts of subpopulations of neurons which express their marker at specific levels. These levels might depend on the regions or the cell types, but it is an opportunity to improve the BBCAv2 locally.

Future analysis will include estimating cell density from the ISH images using automatic point-detection algorithms as presented in Erö et al. [8]. These algorithms will provide new cell counts to rectify our estimates. However, cell counting requires significant resources to be applied to the whole brain, and will also be biased by cell overlapping, especially in highly dense regions such as for Cerebellum, as discussed in Erö et al. [8].

### 3.4.4 Limitations of the assumptions of the pipeline

The BBCAv2 estimates of the different cell densities are based on the five assumptions described in Section 3.2.3. These assumptions come from findings mostly published on the isocortex and represent the rules we used to categorize all cells in the mouse brain [59], [60], [66]–[71]. However, for each of our first four assumptions, we found exceptions reported in literature:

1. A small portion of astrocytes from the visual cortex have been reported to express GAD67 [90].
2. Szabolcsi and Celio [91] observed that ependymal glial cells express PV. These cells are surrounding the ventricles of the mouse brain which regions are present in the AV. No neurons were placed within these regions as can be seen in Figure 3-5; these cells should therefore have a minor impact on our results.
3. Zeisel et al. [66] reported that PV, SST, and VIP populations overlap in most regions of the brainstem. There are many types of cells co-expressing at least two of our markers with different levels of expression and we found no automatic solution to detect each of these cells using the data available, since the ISH data from the AIBS comes from different species. Lee et al. [92] also pointed out, using double ISH staining, that a population of neurons in the isocortex is co-expressing PV and SST.
4. Zeisel [66] also found that PV and SST are expressed in excitatory neurons and VIP in modulatory and non-inhibitory neurons in some regions of the brainstem such as the colliculus and medulla, which were detected as errors in our reports (see Figure 3-5A).

If validated by additional studies of density estimates in the regions where these exceptions have been reported, we could justifiably subtract a proportional amount from the deduced numbers of GABAergic cells in future versions of the model.

As we describe in Section 2.4.1, some of the inner region variation of cell densities can be lost applying the assumption 5. This concerns gradients within the neuronal population since the neuron distribution is derived from the Nissl volume. For instance, in the hippocampus CAs, Jinno et al. [43] have shown a dorsal to ventral distinction in the density of the inhibitory neurons, which is not reflected by the subdivision of the AV. Additional subregions for the AV have to be created to include these findings.

### 3.4.5 Reliability of quantitative studies on inhibitory neurons

One of our objectives was to match regional estimates from literature with our neuron types' density distribution. To do so, we performed a literature review of quantitative studies on PV+, SST+, VIP+, and GAD67+ neurons in the mouse brain (see Section 3.2.4).

Unfortunately, we found that literature does not provide a complete coverage of the whole mouse brain for GAD67. This can be explained by the difficulty to perform cell counting or cell detection on large tissue volumes, or by the fact that some regions are more frequently studied than others (e.g., barrel cortex, hippocampus CAs, see Figure 3-3B). Moreover, GAD67, PV, SST and VIP are extensively used genetic markers in literature, this is not the case for most of the markers available (e.g., LAMP5). This means that the more precise our cell composition becomes the more difficult it will be to gather enough data points from literature and obtain meaningful transfer functions (see Section 3.2.5).

We also found that studies reporting similar measurements of neuron densities sometimes yielded very disparate results (see Figure 3-3B and Table 3-1). Literature density estimates can be biased by the age, sex, staining technique, neuron counting method, or measurement of the volume of the region of interest [15], [16].

Some of the densities used to build our transfer functions were computed according to cell counts from literature divided by the volumes of the regions of the AV, and potentially prone to errors as we pointed out in Section 2.2.3. Some regions are also known to be purely inhibitory (e.g., layer 1 of the isocortex) and we therefore used the Cell Atlas neuron densities, which are themselves subject to errors as discussed previously.

### 3.4.6 Variability of the transfer functions

Fitting a transfer function from region mean intensity to cell type density is a required process to obtain the density of inhibitory cell type in regions where no literature data are available (see Section 3.2.5). However, as shown in Figure 3-4, the gathered points display a great variability which makes it difficult to find a function to represent the relation between region mean intensity and cell density. This variability might come from the ISH data used to fit the model, or from the literature itself (as discussed in the previous sections).

In our future work on the Cell Atlas pipeline, our transfer function will be improved by creating more subregion groups for which separate fittings will be done. These subgroups could correspond to different expected soma sizes for the cell type of interest. However, this would reduce the number of points available to fit the function and might result in an overfit.

We will also test other monotonically increasing functions to fit the cloud of points of Figure 3-4 such as sigmoid functions to integrate any threshold dynamics for low expression values or saturation for high expression values.

Overall, we found that the fitting of the transfer functions from marker expression to cell densities is sensitive to errors linked to the datasets used as input which can explain some of the BBCAv2 results and the fitting low coefficient of determination.

### 3.4.7 Consolidation of neuron density estimates

As we discussed in Section 3.4.5, literature reliability on estimates of cell counts is currently insufficient. Similarly, the fitting of these literature values to marker expression suffers also from the lack of coverage and remaining artifacts of the ISH datasets (see Section 3.4.3). We described a method in Section 3.2.6 to reconcile our disparate estimates into the consistent framework of the BBCAv2. This method aims to minimize the amount of correction needed to obtain a coherent solution.

In our minimization function (Equation 3-11), we chose to divide the differences measured according to the standard deviation of our first estimates ( $\sigma$ ). This parameter accomplishes two purposes: first, it reduces the impact of dense regions on the global minimization score, and second, it gives more weight to estimates where either literature or the fitting are more confident. Other  $\sigma$  values will be tested in the future, such as the total number of neurons in the region to limit the impact of neuron density and in turn, give the same relative weight to each region. Similarly, we will measure the impact of adding an extra weight to maintain the proportion of each cell type within a region. This new constraint will have to be balanced with the previous one.



We assumed that the distribution of neurons in the mouse brain using the BBCAv1 method is correct, which means that whenever there was a contradiction between this distribution and the BBCAv2 inhibitory neuron estimates (coming from literature or the fitting), the error will be attributed to the latter. BBCAv2 results are therefore biased by BBCAv1. The results of Erö et al. have been validated against literature counterparts [8] but might be incorrect in some specific regions (as depicted in Figure 3-11C). The BBCAv2 seems however to be more in agreement with literature than the BBCAv1 (see Figure 3-11DE). In a future work, we will therefore test our novel approach to estimate cell, neuron, and glia densities in the mouse brain (step 2 of the pipeline - see Figure 3-1) based on literature and the fitting. To this end, we will extend the literature review from Keller et al. [17]. The optimization process from step 4 will have to include these new constraints and variables to solve them all at once.

### 3.4.8 Changes in counts of inhibitory neuron in the brain

We obtained a new distribution of inhibitory neurons in the mouse brain, including a new distribution of PV+, SST+, and VIP+ neurons. The ratios of our inhibitory neuron subtypes counts divided by the neuron counts of the BBCAv2 fall near the numbers reported by Kim et al.: 4.03% PV+, 3.29% SST+, 0.63% VIP+ neurons [16]. The differences can be explained by the integration of other sources of literature to estimate these neuron population densities in the brain.

In the BBCAv1, the total number of inhibitory neurons taken from Kim et al. [16] was used to constrain the whole atlas. This resulted in an underestimation of the total number of inhibitory neurons in the brain. Since our new pipeline now also accounts for the Rest population, estimates in BBCAv2 are higher compared to BBCAv1. The three regions with the most pronounced updates of inhibitory neuron densities are: isocortex, olfactory areas, and cerebellum (see Table 3-3).

In the cerebellum, only the granular layer was updated since the molecular layer was already marked as purely inhibitory in both versions of the Cell atlas pipeline. As we observe in Figure 3-11AB, this BBCAv2 model seems to display too many inhibitory neurons in the cerebellar cortex, granular layer and in the thalamus in general when compared to the GAD67 ISH experiment. Some interneurons of these regions seem to co-express PV and SST as pointed out by Zeisel et al. [66], and as can be seen with a manual inspection of ISH slices of PV, SST and GAD67. As we discussed in Section 3.4.4, we can implement this specificity in our assumptions.

In the olfactory areas, we have collected density estimates from literature [93], [94]. These estimates compensate for the poor coverage of the olfactory bulb by the GAD67 marker experiment (see Figure 3-2D) and result in larger inhibitory cell density estimates in BBCAv2.

In the isocortex, the BBCAv2 estimates an overall higher proportion of inhibitory neurons. In some regions, such as the barrel field or the auditory areas (see Figure 3-8B), the ratio of inhibitory neurons in the isocortex seems to be close to the 20% reported in Markram et al. [4] and Meyer et al. [87]. In other regions, such as the motor areas, the ratio of inhibitory neurons averages above 30%. For most of the isocortex regions, we gathered at least one density estimate for PV, SST and VIP expressing neurons from the literature; the ratio of inhibitory neurons in these regions is therefore minimized by the sum of these literature values divided by the counts of neurons from the BBCAv2. We studied the distribution of these ratios in the isocortex and found its average value to be 17.5% and its maximum to be 34.4%. The distribution of the ratios of GAD67 counts from literature divided by the counts of neurons from the BBCAv2 spans from 16.9% to 34.4% with a mean value at 28.5%. This indicates either that the expected 20% ratio of inhibitory neurons is incorrect in some regions of the isocortex or that the distribution of neurons from the BBCAv2 obtained at step 2 is incorrect (see Figure 3-1).

We also found that the Rest population corresponds to a large proportion of inhibitory neurons in the whole mouse brain. In the isocortex, multiple literature findings show that the proportion of the Rest population should be low [60], [92]. However, this is not the case in the other brain regions. For instance, the striatum is a region almost if not entirely inhibitory (see review from Tepper et al. [95]). Among the inhibitory neuron population in this region, 95% of them are medium spiny neurons. This population has been shown by Zeisel et al. [66] to be expressing neither PV, nor SST nor VIP. Similarly, Zeisel et al. reported multiple neuron types belonging to the Rest population, especially in the olfactory bulb and in the brainstem. This indicates, as discussed in Section 3.4.4, that our equations might need to be refined in some regions of the brain, based on literature composition.

Furthermore, the first two steps of the pipeline are based on the approach of Erö et al. [8], i.e., it uses global estimates from Herculano-Houzel et al. [22] and the Nissl volume to estimate cell densities in every region of the brain. As we saw in this study, this approach is sometimes not in agreement with literature due to artifacts in the Nissl dataset (see Figure 3-5A and Figure 2-2) or misalignments of the latter with the AV (see Section 2.2.3).

### 3.4.9 Changes in the me-type distribution for inhibitory neurons

We obtained a new distribution of me-types in the mouse brain and in the isocortex in particular. As we presented in Section 3.3.4, the change of E/I ratio induced by the change of method to estimate inhibitory neuron densities play a significant role in distribution of morphological types in regions of the isocortex (See Figure 3-9). It has to be noted that the mapping from Roussel et al. [63] is performed on datasets from different species (rat and mouse), targeting different regions of the isocortex (somatosensory cortex, lower limb, and primary visual cortex) which might explain some of the distribution changes observed, despite the normalization process to extract me-features. Moreover, only common features were extracted from both datasets which means that the remaining me-features could not be mapped to a molecular identity. These features might however be important to distinguish different populations of neurons in isocortex. Future analyses would have to be carried out in order to weigh the relevance of each feature in the final mapping. Additionally, the same mapping was applied to the entire isocortex, which may explain partially the overall smooth distribution of types across isocortex. However, additional datasets would have to be integrated to improve on this mapping, which might reduce even more the number of common features to extract and map types on.

# Chapter 4 Whole mouse brain connectivity atlas

A first version of the cell-to-cell whole mouse brain connectome was obtained from a previous work in our laboratory [2]. This model was built using two approaches to deal respectively with the long- and short-range connections in the brain. In this chapter, we refine the short-range connectivity by integrating connectivity data from the literature. We also analyze how the refinement of the cellular composition in 0 impacts our connectome of the mouse brain.

## 4.1 Introduction

### 4.1.1 Long range connectivity reconstruction

To model neural connections in the brain, there are two types of information that are required. For each neuron, we need to define its target neurons and how many efferent synapses it creates with these targets. To answer these questions, Erö [2] decomposed the whole mouse brain connectome modeling into two distinct parts: Mesoscale and Microscale connectivity. First, the author categorized the neurons of the cell atlas based on the type of efferent connections they create. The three categories were the short range (for Microscale connectivity), long range (Mesoscale connectivity) and both ranges. The neurons were assigned to the distinct categories based on their neuron/electrical type (see Section 3.2.8), filling the table below.

	Long-range	Short-range	Both
Excitatory neurons	13940990	375363388	8265965
Inhibitory neurons	39833	11921494	0

Table 4-1 Cell atlas neuron counts based on efferent connection types

This table shows the counts of neurons from the BBCAV1 sorted by the type of efferent connections that these neurons create. These connection types are based on the source neuron type as defined in the BBCAV1 (see section 3.2.8). In absence of literature regional neuron type distribution, the excitatory neurons are considered to create both long- and short-range efferent connections, while inhibitory neurons create short range connections. Modulatory neurons are not connected to the circuit as their impact on the neural circuit is still poorly understood.

Then, using these neuron categories, Erö defined the following equations to describe the mean number of efferent synapses each neuron creates in the mouse brain:

$$M_{brain} = m_{out}^{long} \cdot (N_{EL} + N_{EB} + N_{IL} + N_{IB}) + m_{out}^{e,short} \cdot (N_{ES} + N_{EB}) + m_{out}^{i,short} \cdot (N_{IS} + N_{IB}) \quad \text{Equation 4-1}$$

$$M_{brain} = 3.0 * 10^{11}, m_{out}^{e,short} = 2630, m_{out}^{i,short} = 957, \rightarrow m_{out}^{long} = 10118$$

In Equation 4-1,  $M_{brain}$  corresponds to the total number of synapses in the brain,  $N_{EL}$  is the number of excitatory neurons with long range connectivity,  $N_{ES}$  is the number of excitatory neurons with short range connectivity,  $N_{EB}$  the number of excitatory neurons with both long- and short-range connectivity (respectively  $N_{IL}$ ,  $N_{IS}$ ,  $N_{IB}$  for inhibitory neurons).  $m_{out}^{e,short}$  corresponds to the mean number of efferent short-range synapses per excitatory neuron (respectively  $m_{out}^{i,short}$  for inhibitory neuron), which allow for the computation of the mean number of efferent long-range synapses per neuron  $m_{out}^{long}$ . With this equation, we are therefore capable of measuring the mean total number of efferent synapses that neurons create in the mouse brain.

For long-range connectivity, using the extensive rAAV tracer injection experiments from the AIBS, Erö was capable to reconstruct the mesoscale connectivity from Oh et al. [29]. The axons of the neurons injected with the rAAV during these experiments can indeed be highlighted with green fluorescent protein (GFP). The AIBS made the results of these experiments publicly available in the form of GFP expression datasets of  $100\mu m^3$  resolution, aligned to the AIBS average brain (see Figure 2-1C bottom). Erö integrated these datasets in the BBCAV1 which allowed for the extraction of the neurons that were injected with the virus and their targets. Then, the author used an acceptance rejection algorithm to randomly select potential targets based on the GFP expression in the target area

of each experiment. The resulting connectivity matrix shows a great resemblance with the results from Oh et al. [29], as analyzed by Erö. However, the rAAV tracer experiments are not covering the entire brain. In fact, after mirroring the AAV experiments on the other hemisphere, around 20% of the neural population has been covered by at least one experiment (see Figure 4-1A and Table 4-2). Moreover, neurons injected with the same experiment cannot be distinguished due to the resolution of the datasets. Hence, these neurons will have the same predicted targets. Ideally, we would need to reconstruct each neuron morphology to be able to place them directly in our brain volume and extract the right targets. A good estimation of the confidence of our long-range connectivity matrix can therefore be measured using the ratio of the number of AAV experiments per neuron within a region. We performed these measurements for the connectivity matrix obtained from the whole brain model with 1% of its neurons and 1% of its synapses. From this Figure 4-1B, we can see that the Pallidum is the region with the most injections per neuron (1 per 49 neurons) while the cerebellar cortex is poorly covered by the experiments (1 per 11695 neurons). These analyses can help us in determining regions where new connectivity data need to be integrated to improve the long-range connectivity of our model.

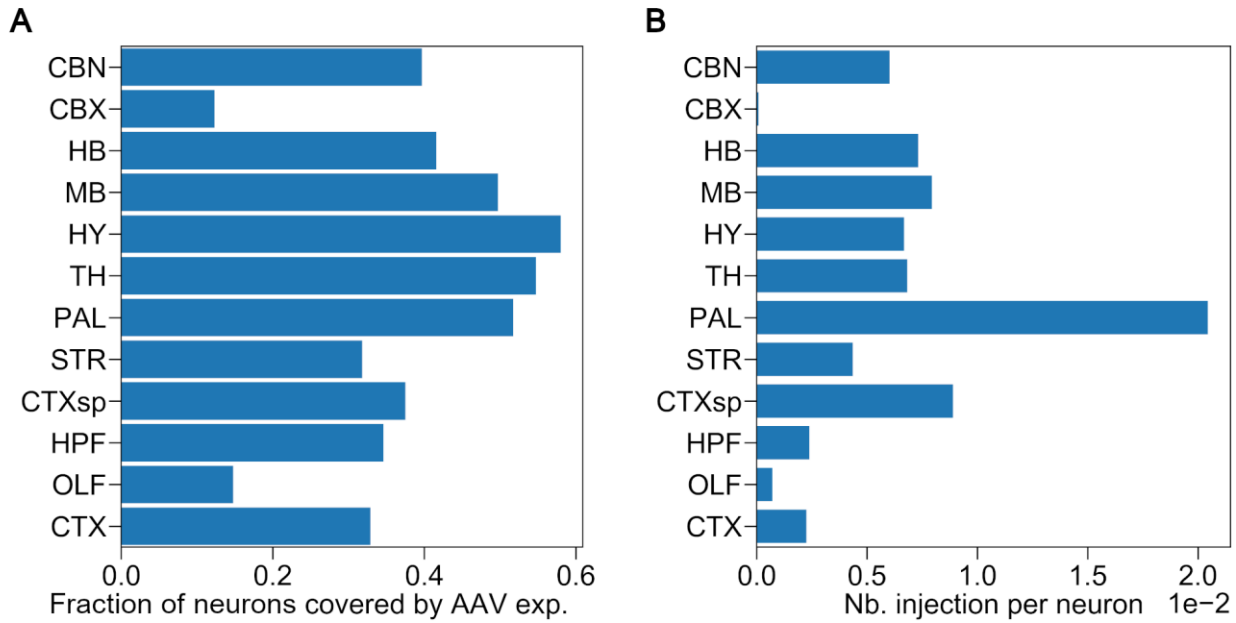


Figure 4-1 AAV tracer injection coverage from Oh et al. [29].

(A) Fraction of neurons covered by at least one rAAV tracer experiment, grouped by brain region.

(B) Average number rAAV tracer experiments per neuron, grouped by brain region.

Region shown: Thalamus (TH), Hippocampal formation (HPF), OLF (Olfactory areas), Cortical subplate (CTXsp), Hypothalamus (HY), Hindbrain (HB), Midbrain (MB), Pallidum (PAL), Cerebellar nuclei (CBN), Cerebellar cortex (CBX), Striatum (STR), Isocortex (CTX).

Region name	Number of Injection	Number of neurons injected	Total number of neurons
Isocortex	202	29474	89644
Olfactory areas	58	12092	81879
Hippocampal formation	117	16968	49035
Cortical subplate	45	1898	5061
Striatum	75	5478	17219
Pallidum	51	1290	2494
Thalamus	93	7457	13630
Hypothalamus	81	7036	12131
Midbrain	103	6455	12983
Hindbrain	77	4371	10512
Cerebellar cortex	36	51876	421035
Cerebellar nuclei	8	527	1328

Table 4-2 Neuron coverage of the AIBS AAV tracer injections.

For each main region of the mouse brain, the number of AAV tracer injection experiments are listed together with the number of neuron injected and the total number of neurons in the region.

### 4.1.2 Short range connectivity reconstruction

For short range connectivity, neurons were assumed to connect to any neuron in their surrounding area. This area was defined as a sphere of diameter  $442.78\mu\text{m}$  surrounding the neuron soma, based on the work of Markram et al. [4]. This radius parameter was the same for every neuron despite the great variability of their morphology as the latter is usually unknown.

To test this model, we compared the m-type to m-type connectivity matrix extracted from Markram et al. [4] in the SSCtx-LL to a similar matrix from our model (see Figure 4-2). The Markram et al. circuit leverages the morphologies of its neurons to create connections between them. Their algorithm detects the apposition of the neuron morphologies which are labeled as potential touches. Connectivity rules are applied to filter correct potential connections. For instance, most of the axon-to-axon apposition will not be considered as potential connections. Then pruning operations are applied to reduce the number of appositions to regroup synapses and match literature distribution of bouton densities. The final connectivity matrix represents therefore a target for our model as it captures morphological detail that our point-neuron model lacks by definition.

We found differences in the distribution of synapses between Markram et al. circuit [4] and ours (see Figure 4-2). First note that our model does not have the same set of m-types as in Markram et al. This is due to a refinement of the excitatory m-types by Kanari et al. [96], which we integrated in our model. This may explain a few differences between the matrices of Figure 4-2. Markram et al. found almost no connections coming from layer 1 neurons and targeting the L6 excitatory populations and conversely connections L2/3 inhibitory neurons send axons to L5 and L6. However, as can be seen in Figure 4-2B, these properties are not matched by our model. Moreover, the ChC population should not target inhibitory neurons even by chance.

A second issue with our model is linked to the presynaptic electrical parameters used to simulate the synapses' dynamics. These parameters are defined based solely on the presynaptic neuron type: excitatory or inhibitory. However, literature reported specific parameter sets to link neuron type together [4], [14]. These electrical parameters are not implemented in our model but play a role during simulation.

It is difficult to improve on the local connectivity algorithm from Erö without placing neuron morphologies and looking at the appositions of axons on dendrites or integrating literature connectivity rules. However, in this chapter, we will implement a recipe-based connectivity where connection types can be refined based on literature findings on the neuron/electrical type of the source and potential neuron targets. This kind of information can be easily implemented to refine the short-range connectivity in numerous regions of the brain.

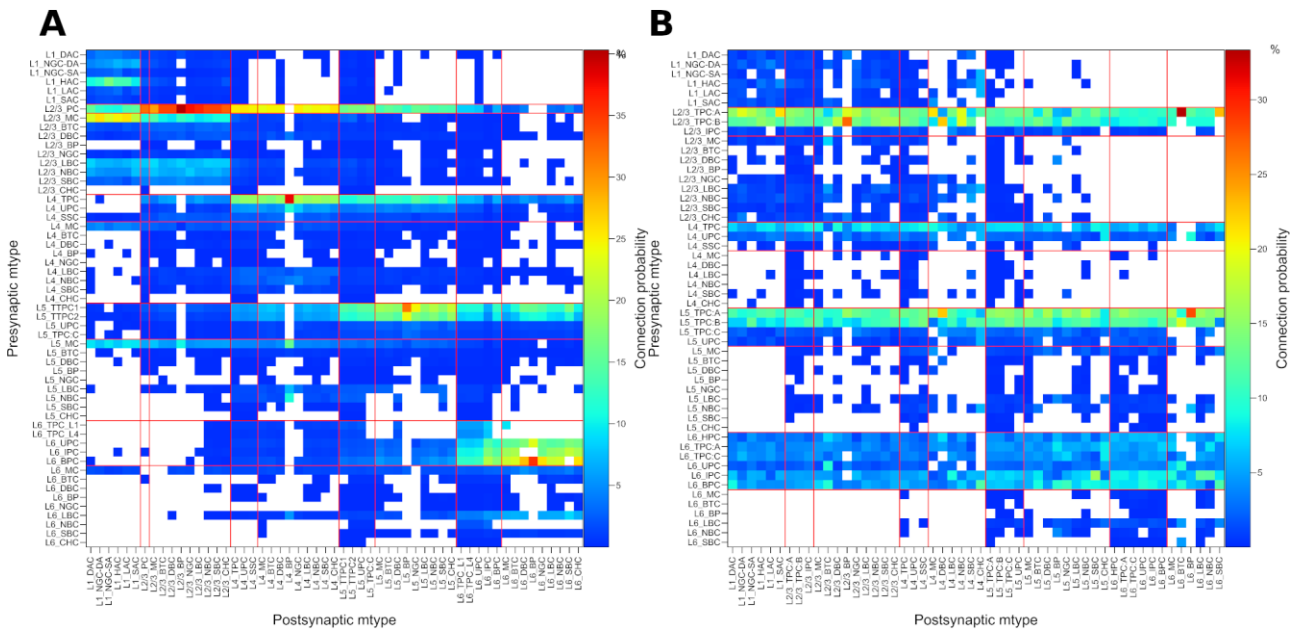


Figure 4-2 Comparison of somatosensory cortex lower limb connectivity matrices.

Matrix of the m-type-to-m-type connection probabilities of the somatosensory cortex lower limb circuit according to the Markram et al. [4] (A) and WBM1 (B). Absence of connections is shown in white. Red lines separate the excitatory from the inhibitory populations. The connection probabilities are normalized according to the size of the postsynaptic population. (A) and (B) use different set of m-types because the labels from Markram et al. [4] were recently refined by Kanari et al. [96].

## 4.2 Methods

### 4.2.1 Literature based refinement of the local connectivity

As we discussed in Section 4.1.1, the connectivity in our point neuron brain model is based on the neuron types defined by our Cell atlas. The refinements implemented in 0 changes the distribution of excitatory and inhibitory neurons, which in turn will change the number of long-range efferent synapses per neuron. Table 4-3 shows the new distribution of neurons in the mouse brain based on their type of efferent connections. We can now estimate the mean number of efferent long-range synapses to be  $m_{out}^{long} = 10923$  based on Equation 4-1.

	Long-range	Short-range	Both
Excitatory neurons	12107819	37529948	7972235
Inhibitory neurons	40143	14054314	0

Table 4-3: **Updated cell atlas neuron counts based on efferent connection types.**

We also want to refine the algorithm from Erö [2] to reconstruct local connectivity. For each neuron that creates efferent short-range connections, this algorithm randomly selects targets within a sphere surrounding the source neuron. However, some neurons target specifically other types of neurons. For instance, the reticular neurons in the thalamus target excitatory neurons from the VPM or POM. Similarly, the electrical parameters of these connections are sometimes even more specific. Yet, numerous studies have been published on the neuron connection types and parameters. This information was not integrated in the pipeline by Erö. We therefore gathered connections and synaptic types from literature through the brain, including isocortex [4], cerebellum [14], and thalamus and hippocampus CAs reconstructions from the Blue Brain Project. We extracted from these papers a neuron type to neuron type matrix of potential connections, with presynaptic electrical parameters for each connection. These datasets were used as recipes during the creation of local connectivity. We used the same spherical kernel to determine potential targets, but we also filtered out connections that were not reported by literature. For instance, Chandelier cells create no efferent connection on other inhibitory neurons, as reported in Markram et al. [4]. Finally, to each connection type was assigned presynaptic parameters based on the same literature sources.

The connection rules extracted from our review are implicit, which means that sometimes information about the location or the types of the neurons was not fully specified. For instance, the rule “PC neurons in any layer of isocortex connects to PC neurons of any layer of isocortex with the parameters x” can be applied in each subregion of isocortex, and to any neurons having this m-type label, no matter his electrical type. To implement these connectivity rules, we had to resolve them for each neuron and for each target neuron that our algorithm found as a potential target. This process is therefore very slow. We could also pre-compute these rules, keeping a list of all potential connections in the brain with their parameters. This technique is however inefficient in terms of disk usage and is difficult to maintain. As we will discuss in the following section, we solved this issue by storing these connection types and their parameters in a database.

### 4.2.2 Create a database of Neuron types

The pipeline to create the whole mouse brain point neuron model leverages the cell types and connection types coming from literature. Especially, as we discussed in the previous Section, connection rules from literature are numerous and need to be explicit to simplify their integration during the process of local connectivity reconstruction. To store and maintain all this information, we constructed a SQL database to store cell types (see database schema on Figure 4-3A), connection types and their point neuron parameters. In this database, a neuron instance has 3 main properties (SQL table neurons):

1. a brain region as defined in the AV,
2. a cell type which here represents a general label used to define the neuron in literature
3. an electrical type (e-type) that defines the electrical behavior of the neuron.

Each e-type can be represented with a point neuron model with a set of parameters. There are many point-neuron models capable of representing each e-type, and each has its own set of variables and constraints. To simplify the integration of distinct types of point-neuron models, the table storing each point-neuron model electrical parameters (SQL table neuron\_models\_params) was separated from the neuron instance (SQL table neuron\_instances\_constants). This implementation of the database also allows for the integration of distribution of parameters and keeps track of their units (respectively SQL table distributions and units). We used this

database to store the distribution of neurons collected from literature [4], [14], [19]. We leveraged this database during the process of assigning me-types and their point-neuron parameters to the neurons of the WBM.

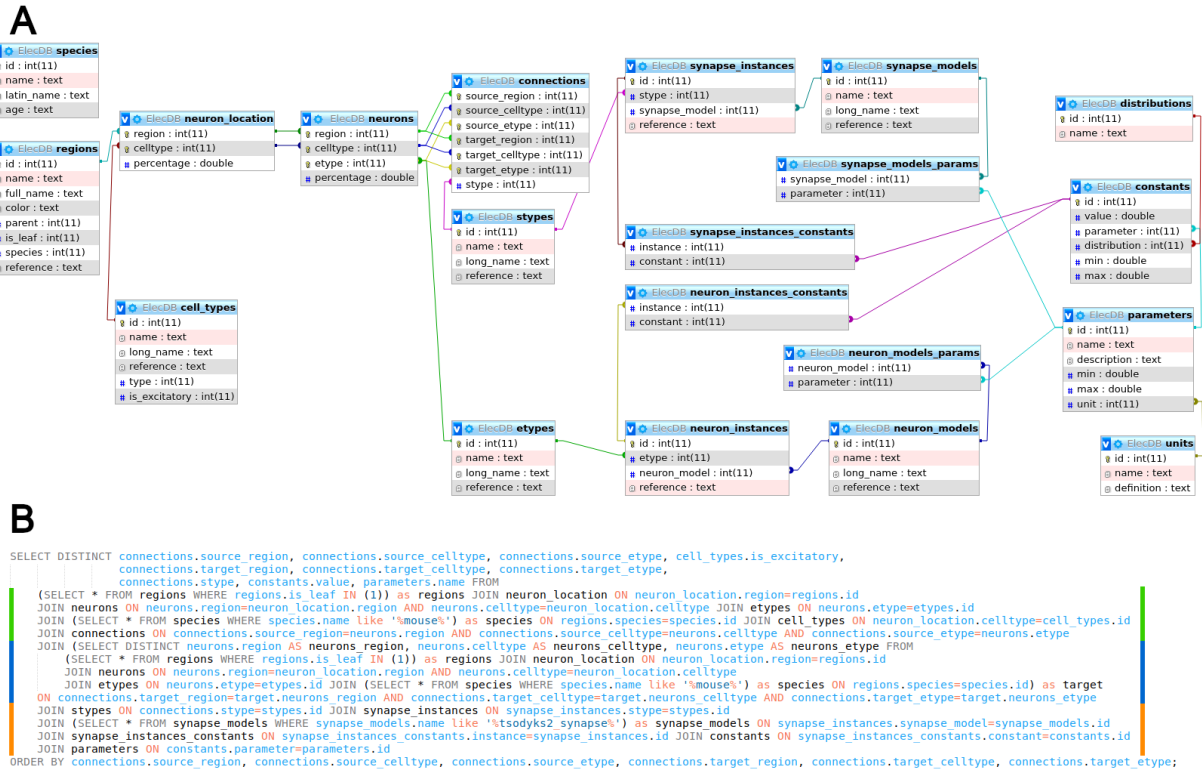


Figure 4-3: Database schema for a point-neuron network database.

(A) SQL schema of the database used in our pipeline to store neuron types, connection types and their parameters. SQL tables are represented with boxes with each of their parameters listed below. Primary key parameters are annotated with a key symbol. The tables are connected with foreign key constraints represented as colored lines linking the boxes.

(B) Example of a SQL query to extract the synaptic parameters for every neuron of the mouse brain. The query is composed of three subqueries as indicated by the color bars. The first subquery (green) extracts the presynaptic neurons. Next, the postsynaptic neurons are queried (blue). Finally, the parameters for the Tsodyks-Markram synapse model are selected, in accordance with the pre- and postsynaptic neuron types (orange).

We also leverage this database to store the connection types and their electrical parameters. A connection (SQL table connections) links a source and a target neuron type together with a synapse type (S-type). Each S-type can be represented with a presynaptic synapse model with a set of parameters. As for the neuron instances, we allow for the integration of multiple synapse models.

We used this part of the database to store the connection rules described in the previous Section. To do so, we had to make these rules explicit by fetching every neuron type matching the description of the connection rule. However, the great number of brain regions, neuron type and electrical type can make the number of potential connections very big. In particular, the isocortex has 82 different me-types of neurons, which are distributed in 426 leaf regions of the brain. Even taking into account their layer distribution, if all these neurons could connect to any other neuron of the isocortex, the database would have to store  $2^{18440} \approx 9.84 \cdot 10^{21}$  entries. Most of these entries cannot appear in the brain due to the distance between brain regions. For instance, no short connections can be made between the Frontal cortex and the Visual areas. For each isocortex subregion, we therefore computed a list of potential target regions within the local connectivity distance range. To do so, we reused the shading algorithm used for orientation computation (see Section 2.2.4), to extract all voxels surrounding the source region and within the radius of the sphere kernel used for local connectivity. This limited the list of potential target regions for each source region and in terms the total number of entries in the database. This part of the database was leveraged during the creation of local connectivity in the brain to implement results from literature.

## 4.3 Results

### 4.3.1 Refinement of the cell-to-cell connectivity matrix of the whole mouse brain

We have defined a new method to refine the cell-to-cell connectivity matrix in certain regions of the mouse brain based on literature findings. This method leverages a new database that stores cell and connection types as well as their electrical parameters.

Sup. Figure 10 shows the new connectivity matrix of the whole mouse brain. The connectivity matrix was impacted by the changes in the neuron composition of the brain that we presented in Chapter 3. To analyze the impact of our method on connectivity, we will focus on the isocortex and more specifically on the somatosensory cortex lower limb (SSCtx-LL). The m-type-to-m-type connectivity matrix of the SSCtx-LL is shown in Figure 4-4. We note that the connections from Chandelier cells are no more targeting the inhibitory neuron population, indicating that our recipe-based connectivity algorithm is working properly. Compared to our previous version of the connectivity, we observe no noticeable difference in the connection probabilities linked to the inhibitory neuron types. However, the efferent connections of the excitatory population are more evenly distributed which is the result of the new distribution of neuron types in the brain (see Figure 3-9C). Our local connectivity algorithm indeed randomly samples neurons in the vicinity of the source neurons to create connections. This connectivity matrix therefore highlights distribution of neuron m-types in the vicinity of each m-type. However, the distance that defines the vicinity of a cell might not be adapted when we compare our connectivity matrix to the one from Markram et al. [4] (see Figure 4-2A). For instance, the L6 excitatory population seems to send very few axons to the L1 and L2. To take this knowledge into account we would need to know the orientation and length of the axon of each m-type in the region.

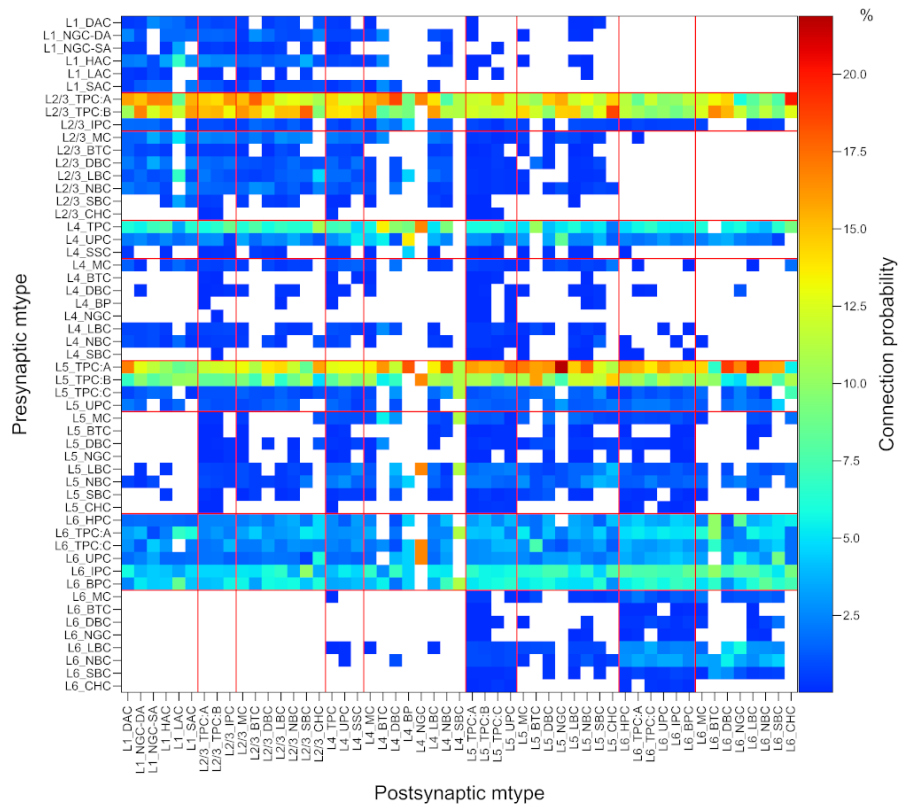
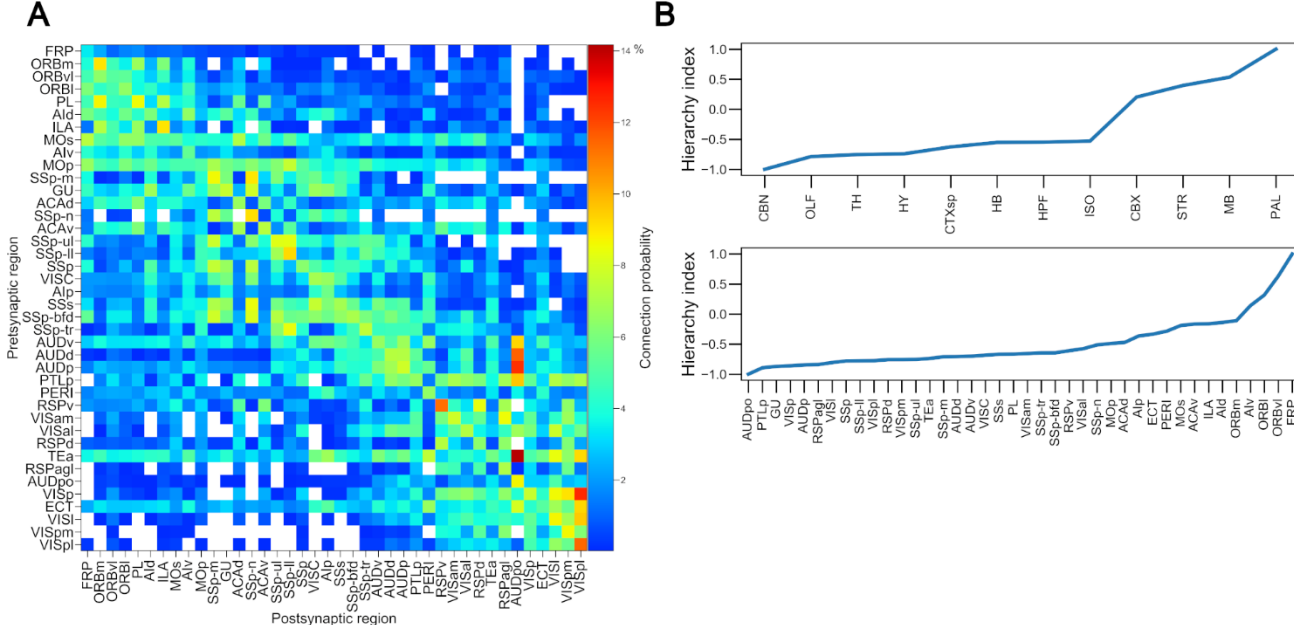


Figure 4-4: **Generated SSCtx-LL connectivity matrices of the refined whole mouse brain model.**

Matrix of the m-type-to-m-type connection probabilities of the somatosensory cortex lower limb circuit according to the WBM2. Absence of connections is shown in white. Red lines separate the excitatory from the inhibitory populations. The connection probabilities are normalized according to the size of the postsynaptic population.

We also analyzed the inter-region connectivity within the isocortex. The short-range connectivity probabilities in-between brain regions are inversely proportional to the distance between these regions and as can be seen on Figure 4-5A, this property is clearly visible in the matrix when regions are sorted according to their position along the rostral-caudal axis. We then investigated the contribution of each region in the isocortex based on its long-range connections. This was done by computing the hierarchy index of





(A) Matrix of the region-to-region connection probabilities of the Isocortex. Absence of connections are shown in white. The connection probabilities

(B) Hierarchy index of the subregions of the brain (top) and of the subregions of the isocortex (bottom).

We have created an SQL database to store information about the local composition of the brain region in terms of cell types, but also

We have created an SQL database to store information about the local composition of the brain region in terms of cell types, but also the electrical parameters of these cell types which are associated with a point neuron model. We were also able to integrate properties of different connection types and their synaptic parameters. This new database is used in the whole brain model workflow to refine the cell types of the cell atlas with me-types distributions from the literature (step 5 of Figure 1-1, see also Section 3.2.8). Connection types can also be assigned to the short-range connections obtained from our algorithm whenever a literature source is available (step 7 of Figure 1-1, see also Section 4.2.1).

This database can store the information about the different types of neurons and synapses and their parameters, which makes it very useful to build any kind of point-neuron network, from various species. For instance, we were able to store the parameters of the Potjans and Diesmann network [97] in our database and rebuild it. We managed to reproduce Figure 6 from Potjans and Diesmann (see Figure 4-6). Our database could therefore also be used more generally by the scientific community to store point-network parameters and replicate experiments performed on point-neuron networks.

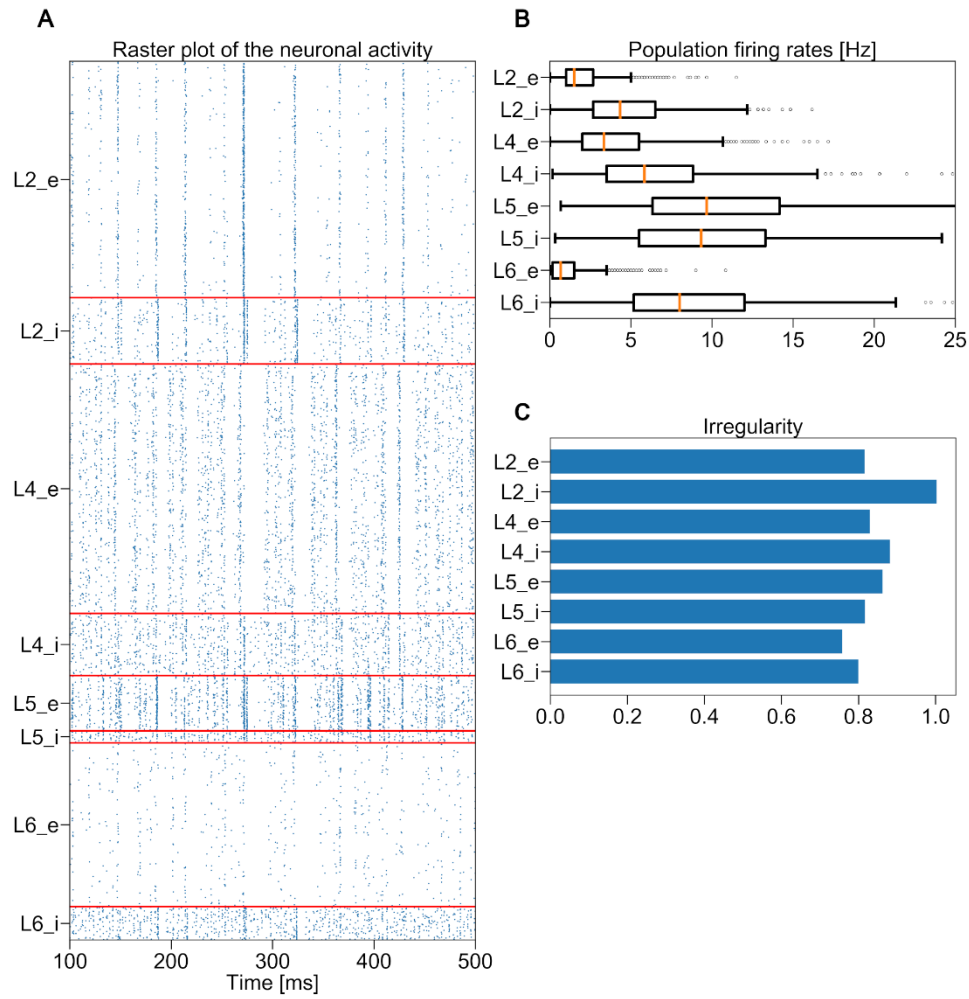


Figure 4-6: **Simulated spontaneous cell-type specific activity.**

Results of network activity reproduced from the work of Potjans and Diesmann [97].

(A) Raster plot of spiking activity recorded for 400 ms of biological time of layers 2/3, 4, 5, and 6. Their excitatory (\_e) and inhibitory (\_i) populations are separated by red lines.

(B–C) Statistics of the spiking activity of all 8 populations in the network based on 1000 spike trains recorded for 60 s.

(B) Boxplot of single unit firing rates. Crosses show outliers, orange lines indicate the mean firing rate of the population.

(C) Irregularity of single-unit spike trains quantified by the coefficient of variation of the interspike intervals.

However, as can be seen in Figure 4-3A, the database has a complex structure which makes it difficult to construct SQL requests. For instance, using this architecture the SQL request to fetch the synaptic parameters of the Tsodyks-Markram synapses between any neuron of the mouse brain is shown in Figure 4-3B. This complexity can prevent the diffusion of this database to the community. We therefore implemented a python interface which simplifies the SQL request creation for fetching or storing information in the database.

## 4.4 Discussion

### 4.4.1 Impact of the new cell atlas on the whole brain connectivity matrix

The changes of method introduced in Chapter 3 produced a new version of the cell atlas (BBCAv2) which contains overall more inhibitory neurons (see Figure 3-7 and Table 3-3) compared to our previous version. This translates in our connectivity matrix with more local inhibitory connections (+18% see Table 4-1 and Table 4-3). We can predict that this change will induce a stronger and faster inhibition of any strong activity in the brain in simulations of the resulting point neuron network. In the isocortex, the new distribution of m-types obtained with Roussel et al.'s pipeline [63] (see Section 3.3.4) distributes more evenly the m-types than the original distribution from Markram et al. [4]. In our connectivity matrix, this translates in the afferent connections of each m-type to also be more evenly distributed (comparing Figure 4-2B and Figure 4-4). Note that our analyses were conducted on a single instance of a scaled version of the SSCTX-LL for both versions of the whole brain model (1% of the neurons and 1% of its synapses) and that the local connections with the surrounding cortical regions are not considered. More the local connections are chosen randomly in a sphere surrounding the source neuron. These two properties explain why some of the neuron types have few connections with the rest of the circuit. Further statistics on the counts of connections should therefore be conducted with a full scaled version of the whole brain model.

### 4.4.2 Cell-type recipe-based connectome of the brain

We introduced connection types in our whole brain model, that we linked to electrical parameters gathered from the literature. This connection types were associated with a source and a target neuron type that were assigned to neurons of our cell atlas. This method is a clear improvement from our previous version of the whole brain model and facilitate the integration of literature findings on regional connectivity. We could indeed implement connectivity rules that prevent some connections to be created in our model. For instance, we forced local connections from cortical neuron to be limited to target only subregions of the isocortex. This method could be extended to integrate long-range connectivity rules from literature. For instance, the medium spiny neurons of the striatum can be categorized in two types, each of which target its specific region of the brain: the globus pallidus externa or the substantia nigra reticulata [19]. However, literature findings might contradict the rAAV tracer injections used to create the long-range connectivity. Additionally, connectivity rules from the literature are very specific since they require the identification of the neuron types and a close inspection of their synapses. That is why, long range connectivity rules from literature are very sparse and usually do not cover the full population of neurons present in the region of interest. Our connectivity algorithm also does not consider distance- or type-based connection probabilities or counts. For instance, globular bushy cells of the cochlear nucleus create a single synapse to neurons of the medial nucleus of the trapezoid body called the calyx of Held [98]. This integration of these properties should be the subject of a future version of our connectivity atlas.

To improve our local connectivity, we can also consider neuron orientations and morphology properties. For instance, most cortical neurons orient their dendrites to the upper layer of the cortex and send their axon to deeper layers of the cortex. Based on this information and the soma locations of each neuron in the cortex, we could therefore limit the efferent connections of a source neuron to certain parts of its surroundings. This technique has been implemented in Reimann et al. [99] using neuron average morphologies and the authors have proven this method to be able to reduce drastically the number of potential connections in a reconstruction of a column of the rat cortex. However, if morphologies are provided for a portion of the mouse brain, it might be preferable to fully reconstruct the local connectome using our cell atlas model and the method presented in Markram et al. [4]. Once the detailed circuit have been reconstructed, we can use simplification methods, such as Rössert et al.'s [100], to capture the connectome and its electrical properties from the detailed circuit and convert it back into a point-neuron network.

### 4.4.3 Storing connectivity rules and electrical parameters into a database

We have constructed a database to store electrical parameters and properties linked to neuron types and their connections, based on literature findings. This includes distribution of neuron types as well as point-neuron model parameters, and synapse model parameters. With this database we could store almost all relevant data used to build any point-neuron network from the literature. As we discussed in the previous Section, at the moment, we are not integrating connection probabilities, but this could be done in a future version of our model. We also created a python interface to facilitate the creation of SQL queries to fetch and add data to the database. We want here to make this tool available to the neuroscience community to facilitate the access and integration of data on neuronal connectivity. The database can further be extended to integrate neuron or synaptic properties that we are not considering in our whole mouse brain modeling effort.

We leverage this database in our pipeline to build the whole mouse brain model to refine the neuron type and their point neuron parameters in regions covered by literature. Similarly, we assigned connection types and synaptic properties to the short-range connections obtained through our method. To do so, we transform each connectivity rule from the literature into a list of potential connection types in the brain. When building the local connectivity in the mouse brain, we can directly use this list without having to resolve each connectivity rule again. This solution implies however to save each potential connection of the brain, which means potentially to save any pair of neuron instances in the brain. While most of these connections are not biologically possible, the total number of potential connections might be substantial. For instance, our database stores thousands of cortical connection types but these do not include the long-range connection types linking these regions together. By combination, the size of the database scales rapidly with the introduction of any new brain region, any neuron type, or any electrical type. This in turns increase the size of the database disk usage but also the time to process SQL requests on the connections. A solution will be to decompose the database into smaller files based for instance on brain regions group.

# Chapter 5 Simulation of the whole mouse brain model

## 5.1 Introduction

The whole- mouse brain (WMB) model is a point neuron network that can be simulated to replicate biological experiments. Results of these simulations will also be used as benchmarks to drive new requirements to our pipeline in a cycle of generations/refinements.

Erö was able to perform a whole brain simulation of the resulting point-neuron model using the NEST simulator [2]. We will also use NEST for its capacity to scale with the size of the network to simulate and the available computational resources [101], and its wide range of point neuron models already implemented.

For our benchmark purposes, we will focus on the somatosensory cortex, lower limb (SSCTx-LL) for which we have simulation results from Markram et al. [4] on the rat. We will perform three types of simulation of this region:

- The first experiment will study the SSCTx-LL isolated from the rest of the brain. Here, each neuron will be stimulated by independent Poisson process generators. The purpose of this simulation is to measure the activity of the SSCTx-LL at a resting state, over a prolonged period. We want indeed to perform statistical analyses on the spiking activity of each neuron population, which require a long period of stable activity.
- The second experiment will focus on activity triggered by the Thalamus VPM which is one of the major external inputs of this region [4] (see Sup. Figure 8A). Here we would like to test the capacity of our model to reproduce findings from the literature on Somatosensory cortex activity when the equilibrium of the steady state activity is disrupted [4], [102].
- Finally, we will analyze the activity of the somatosensory cortex in the context of a whole brain, to see how the activity propagates to other regions and how the activity of other regions impacts the somatosensory cortex.

We obtained a new version of the point-neuron whole mouse brain model with the refinement of the cell types as described in 0, as well as the local connectivity refinements of Chapter 4. We will benchmark this new version using the same experiments described previously.

As described in Erö [2], the full-scale version of the WMB model requires a lot of disk space, as the synaptic parameters alone take up more than 10Tb. Furthermore, for any simulation of the model, the time to simulate the network and the size of data generated will scale with its duration. While Erö [2] was able to generate and simulate a full-scale version of the WMB model, this simulation requires too many resources to be performed for benchmarking purposes. However, we should keep in mind that the dynamics of a scaled version of the network might be different from its full-scale counterpart.

## 5.2 Methods

### 5.2.1 Point neuron network parameters

#### 5.2.1.1 Point neuron model

We choose the Adaptive exponential leak integrate and fire model [103] to simulate point neuron soma activity. This model of point neuron was chosen, as its soma membrane potential equations include adaptation and dynamic spike threshold dynamics and the paper from Naud et al. [104] provides a full set of parameters for major electrical types of the brain. In absence of an electrical type defined by regional literature distribution (see Section 3.2.8), we chose the continuous accommodating (cAC) electrical type for excitatory neurons and the continuous non-accommodating (cNAC) electrical type for inhibitory neurons.

$$C \frac{dV}{dt} = -g_L (V - E_L) + g_L \cdot \Delta_T \cdot \exp\left(\frac{V - V_T}{\Delta_T}\right) - g_e \cdot (V - E_e) - g_i \cdot (V - E_i) - w \quad \text{Equation 5-1}$$

$$\tau_w \frac{dw}{dt} = a \cdot (V - E_L) - w \quad \text{and after a spike,} \quad w = w + b \quad \text{Equation 5-2}$$

Where  $V$  is the membrane potential,  $C$  is the membrane capacitance,  $g_L$  is the leak conductance,  $g_e$  and  $g_i$  are the excitatory and inhibitory conductance and  $E_L$  is the leak reversal potential. The slope factor  $\Delta_T$  is used to drive higher the membrane potential above a spiking threshold  $V_T$  to trigger spikes. The adaptation current  $w$  is triggered by spikes, incremented by the parameter  $b$ , and decaying exponentially with the subthreshold parameter  $a$  and time constant  $t_w$ .

### 5.2.1.2 Synapse model

For the synapses, we chose the deterministic Tsodyks-Markram model [105] (implemented in NEST), as this model can capture short-term synaptic dynamics which play a significant role to regulate brain activity. The presynaptic types were set to the excitatory and inhibitory depressing synapse type from Markram et al. [4].

$$\begin{aligned} \frac{dx}{dt} &= \frac{(1-x)}{\tau_{rec}} - u \cdot x \cdot \delta(t - t_{sp}) \\ \frac{du}{dt} &= -\frac{u}{\tau_{fac}} + u_0 \cdot (1-u) \cdot \delta(t - t_{sp}) \end{aligned} \quad \text{Equation 5-3}$$

Where  $x$  is the amount of synaptic resources available for transmission which decrease with each spike time  $t_{sp}$  but replenish over time with recovery time constant  $\tau_{rec}$ . Facilitation dynamics are defined with the parameter  $u$  which increases with every spike, and the facilitation time constant  $\tau_{fac}$ . For the postsynaptic parameters, we chose the beta-exponential dynamics to model AMPA/NMDA and GABAA/GABAB postsynaptic receptor dynamics. Each receptor has its own set of parameters, as defined in Erö [2]. The delay required for a spike to travel from the presynaptic neuron to its postsynaptic target can be processed for long range connections based on the distance between the two neurons divided by the speed of propagation of a spike in myelinated axons. The latter has been estimated to 840  $\mu\text{m/ms}$  in Erö to match delays of Simmons and Pearlman [106]. The average conductance of the synapses was set according to the distributions of Markram et al. [4], i.e.,  $0.85 \pm 0.44$  nS for excitatory synapses and  $0.84 \pm 0.29$  nS for inhibitory synapses.

## 5.2.2 Resting state activity of the SSCtx-LL

The resting state activity of a brain region is the activity produced by neurons within this region without any external perturbation. This activity is usually linked to random spontaneous release of vesicles producing minis, or more generally to small concentration of neurotransmitters in the extracellular space. These events are sparse but generate small postsynaptic potentials (PSPs). Usually, they can be modeled using random processes on the presynaptic neurons. Our point neuron network uses the deterministic model of Tsodyks-Markram [105] to calculate the presynaptic dynamics. We had to represent these events on neurons as random spikes coming from independent Poisson processes. We used the model implemented in NEST, with a low rate of 0.5Hz, and connected each process with excitatory synapses (with mean conductance weight). These processes were sufficient to produce stable activity within the region. We will study the resting state activity of the SSCtx-LL network with 1% of its neurons and 1% of the synapses. This corresponds to a point neuron network with 1907 neurons connected with 110334 synapses for WBM1; 1905 neurons and 110356 synapses for WBM2. We additionally scaled down the weight of the synaptic connections until the mean firing rate of the neuron population (see Equation 5-4) was below 10Hz. This condition guaranteed a stable spiking activity without any burst. The network was hence simulated for 10 seconds after a warming up period of 500ms, and spikes of each neuron were recorded for analyses. Some neurons did not spike during the simulation. These neurons might be disconnected from the network due to the randomness of the neuron and synapse selection due to the scaling. Hence, they were discarded from the neuron pool during analyses.

First, we measured the firing rate of neurons, which is defined as the expected count of spikes per unit of time. Spikes are discrete random events; hence their underlying firing rate cannot be directly measured. We can however estimate it, convoluting the spike trains with a kernel function  $K(t)$  [107] (see equation Equation 5-4). Here, we used the triangle function as in Nawrot et al. [107] with a kernel width  $\sigma$  equal to 50ms. We reconstructed the firing rate  $Fr$  for each neuron based on their spike train  $(t_1, t_2, \dots, t_n)$ .

$$\begin{aligned} Fr(t) &= \sum_{i=1}^n K(t - t_i) \\ K(T, \sigma) &= \begin{cases} \frac{1}{6\sigma^2} \cdot (\sqrt{6}\sigma - |t|) & t \in [-\sqrt{6}\sigma, \sqrt{6}\sigma] \\ 0 & \text{else} \end{cases} \end{aligned} \quad \text{Equation 5-4}$$

Next, we computed the inter-spikes intervals (ISIs), which is the time between two consecutive spikes from the same neuron (for  $n$  spikes  $X = \{X_1, X_2, \dots, X_{n-1}\}$ ). We can measure the variability of the ISI, computing the square value of their coefficient of variation. As the coefficient of variation of the ISI can be affected by the changes of firing rate, we use here the local approximation CV2 as defined in Holt et al. [108]:

$$CV2 = \frac{1}{n-1} \sum_{i=1}^{n-1} 2 \cdot \frac{|X_{i+1} - X_i|}{X_{i+1} + X_i} \quad \text{Equation 5-5}$$

We can then measure the variability of the spike counts  $N$  using the Fano Factor (FF) [109], defined as the ratio of the var. This factor is usually computed to measure the trial-by-trial variability. We use it here to measure the inter-population variability.

$$FF = \frac{Var[N]}{E[N]} \quad \text{Equation 5-6}$$

Where VAR and E are respectively the variance and mean functions.

If the activity within the region under this state is equivalent to a renewal process, then the Fano Factor should tend towards the CV2<sup>2</sup> value given enough simulation time.

$$\lim_{t \rightarrow +\infty} FF = CV2^2 \quad \text{Equation 5-7}$$

Spike-triggered adaptation dynamics of the AdEx model (see Equation 5-1) can introduce interval correlations of the spike trains [110]. This in turn may alter the Fano Factor. To test this hypothesis, we used Cox and Lewis equation [111] which integrates the linear spike interval correlation factors  $\xi_j$  in Equation 5-7 as follows:

$$\lim_{t \rightarrow +\infty} FF(t) = CV_{corr} = CV2_{\infty}^2 \cdot \left( 1 + 2 \sum_{j=1}^{\infty} \xi_j \right) \quad \text{Equation 5-8}$$

$$\xi_j = \frac{\sum_{i=1}^{n-1} (X_{i+j} \cdot X_i) - E[X]^2}{(n-2) \cdot VAR[X]}$$

### 5.2.3 Response to Stimulus in SSctx-LL

We will perform an experiment where we stimulate the SSctx-LL with neurons from the ventral posteromedial nucleus of the thalamus (VPM). Using the same configuration as in Section 5.2.2, we are now adding the 915 neurons from the WBM1 thalamus VPM (914 neurons for the WBM2) which are represented as Poisson processes. These neurons are all excitatory and connect to the SSctx-LL network adding 18486 synapses for WBM1 (128820 synapses in total); in WBM2 the VPM added 15964 synapses for a total of 126320 synapses. In this experiment, the VPM will send a wave of spikes of frequency 15.3 Hz for 1 s to the cortex and after 500 ms of steady state, mimicking the thalamus activity upon the stimulus of the mouse limbs [102]. The network was simulated for 4 seconds after 500ms of warm up, and spikes of each neuron were recorded for analyses.

### 5.2.4 Whole mouse brain simulation at resting state

The last experiment that we performed targeted to analyze the activity of the somatosensory cortex in the context of a whole brain. Unfortunately, it is difficult to control the overall activity of the brain since we do not know how each subregion contributes to this activity. Here again we will use random point processes to induce activity in each region of the brain. A random process producing spike events of frequency 0.5 Hz was attached to each neuron in the WBM. The synaptic weights were scaled according to the factor found in Section 5.2.2, which was sufficient to produce activity in each of the brain regions. We simulated the WBM network with 1% of its neurons and 1% of the synapses. This corresponds to a point neuron network with 717599 neurons connected with 60958986 synapses for WBM1; 717599 neurons and 58112694 synapses for WBM2. The circuit was simulated for 4 seconds after 500 ms of warm up and the spike events of each neuron were recorded for analyses.

## 5.3 Results

We want to compare the activity of the previous version of the WMB model (WMB1) to the activity of our new version of the network (WBM2) with our refined cell atlas and connectivity rules. For each of the experiments presented in Section 5.1, we will therefore simulate the two versions using the same simulation parameters to assess their differences.



### 5.3.1 Resting state activity of the somatosensory cortex lower limb

We simulated the somatosensory cortex lower limb (SSCtx-LL) isolated from the rest of the brain for 10 seconds, stimulated by a low background noise activity (see Section 5.2.2). The network in this configuration is spontaneously active and stable. Figure 5-1 shows the results of the resting state experiment in the SSCtx-LL for the WBM1 and WBM2. The raster plot activity (see Figure 5-1A) as well as the measure of the firing rate (see Figure 5-1B) confirm that the two networks remain in a steady state activity. First, we observe that the firing rate activity of the inhibitory population is overall more important in the WBM2 compared to WBM1, and conversely the firing rate of the excitatory population is lower in WBM2 than in WBM1. These results were expected since the proportion of inhibitory neurons increased in this region with the refinement of the cell atlas (see Section 3.3.4). Second, the distribution of the Inter-spike intervals (ISI) (see Figure 5-1C) are close to an exponential decaying distribution. This would indicate that each population can be represented as a renewal process. We further investigated the firing rates of each e-type population as their set of electrical parameters might result in different activity patterns (see Sup. Figure 6). The difference of the mean firing rate of these populations between the two versions of the WBM seems to be also related to the change of proportion of each population. Any change of the cell atlas has therefore a noticeable impact on the activity of the resulting whole brain model.

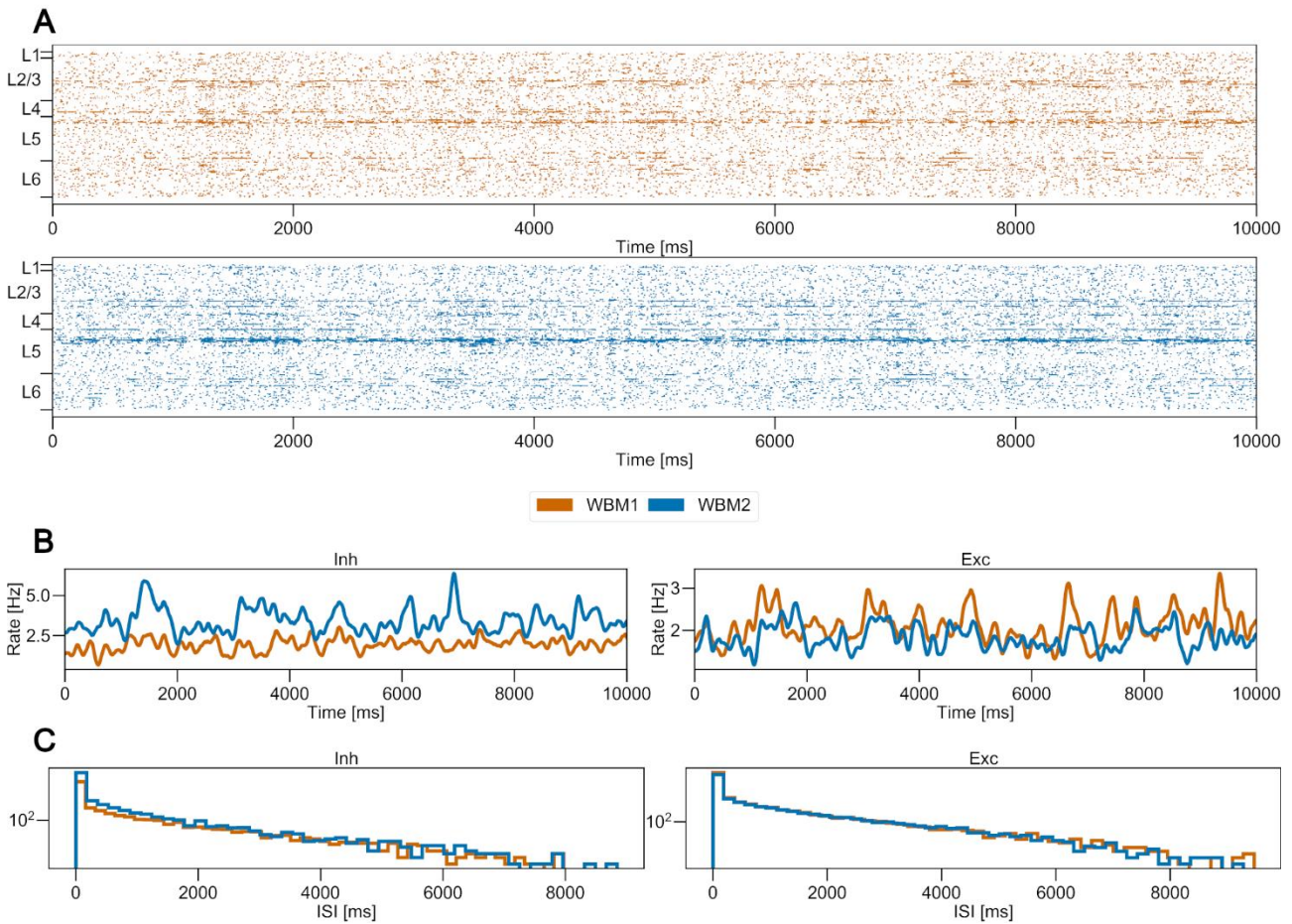


Figure 5-1: **Resting state spiking activity comparison in SSCtx-LL between the WBM1 and WBM2.**

Comparison of the spiking activity between the different versions of the whole mouse brain model in a resting state. The WBM1 (in orange) corresponds to the version presented in Erö [2], while the WBM2 (in blue) represents the refined version after the extension to the cell atlas (see Chapter 30) and the local connectivity refinement of Chapter 4.

(A) Raster plot of the spiking activity in SSCtx-LL, sorted by layer. Neurons are distributed along the y-axis and each spike time is represented as a colored dot along the x-axis. The boundaries between layers are shown on the left.

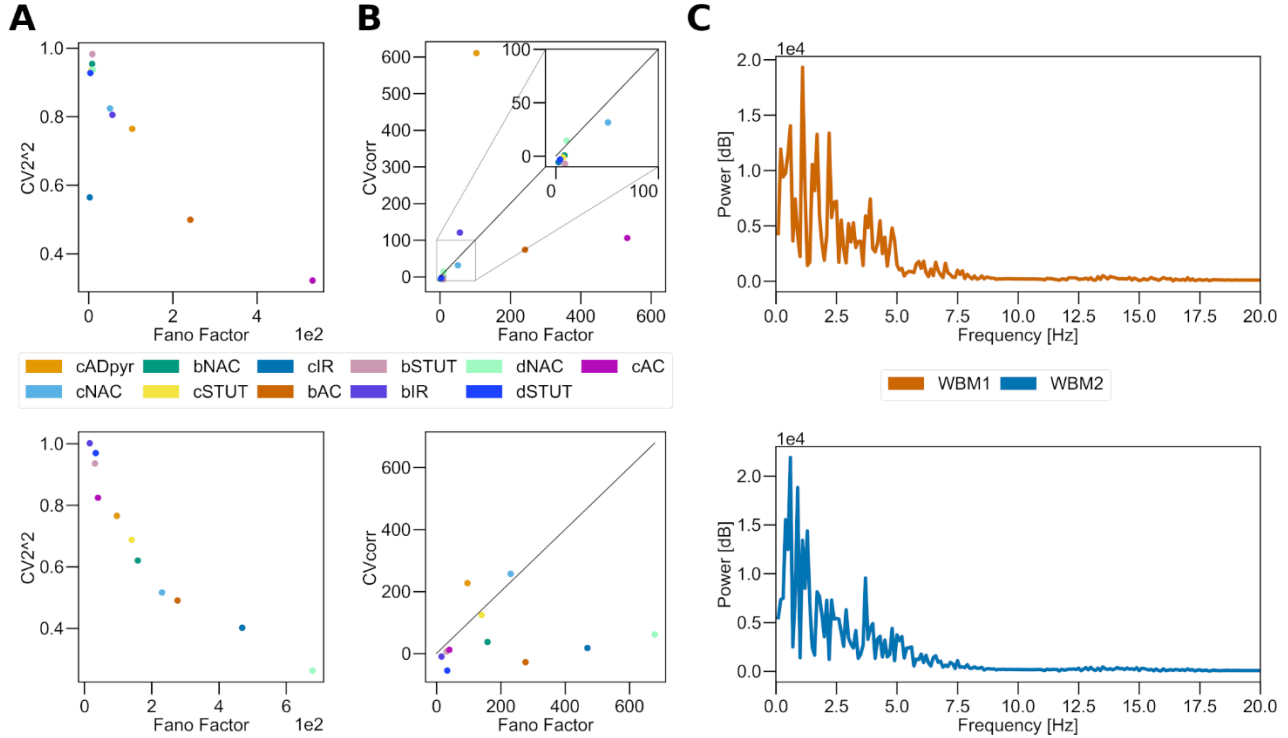
(B) Mean instantaneous firing rate of the inhibitory and excitatory populations of the SSCtx-LL over time.

(C) Inter-spike intervals (ISI) histogram of the inhibitory and excitatory populations of SSCtx-LL in logarithmic scale.

We want to find if the neuronal activity of the SSCtx-LL at resting state can be interpreted as a renewal process. Any deviation from this null hypothesis will trigger more analysis. We will test this hypothesis by studying the variability of the spike rate or also called firing rate, and the inter-spike intervals.



We measured the variability of the ISI distribution ( $CV^2$ ), and the Fano Factor of the spike counts within each population. We then plotted the  $CV^2$  according to the Fano Factor for each e-type population (see Figure 5-2A). First, we note that the Fano Factors for both versions of the model are above 1 which indicates that the activity in each population is more variable than a Poisson distribution [112].



**Figure 5-2 Comparison of the variability of the spiking activity of the SSCtx-LL at resting state.**

Comparison of the variability of the spiking activity in the SSCtx-LL at resting state between WBM1 (top row) and WBM2 (bottom row).

(A) Square value of the coefficient of variation ( $CV^2$ ) of the inter-spike intervals (ISI) as defined in Equation 5-5 over the Fano Factor of the spike counts (Equation 5-6), for each e-type population of the SSCtx-LL.

(B) Corrected coefficient of variation ( $CV_{corr}$ ) as defined in Equation 5-8 over the Fano Factor of the spike counts for each e-type population of the SSCtx-LL. Points falling on the diagonal correspond to a perfect renewal process.

(C) Power spectrum of the SSCtx-LL firing rate for the entire population.

We calculated these correlation factors and plotted the Fano Factors (FF) against the corrected coefficient of variation ( $CV_{corr}$ ) (see Figure 5-2B). Despite the corrections, the FFs are greater than the  $CV_{corr}$  for most of the inhibitory populations for both versions of the WBM which indicates that these populations display very irregular spiking activity compared to a Poisson process. In contrast, the excitatory population (cADpyr) displays a more regular activity. We note that the Fano factors of most of the inhibitory populations are greater in WBM2 compared to WBM1. This can again be explained by the larger proportion of these populations in the new version of the network.

We conclude that the SSCtx activity at resting state cannot be considered as a renewal process for any of its e-type population. This indicates the spike times of different neurons of the SSCtx-LL correlates. The spike correlations can be linked to periodic changes in firing rates. We can detect these oscillations, performing a spectral analysis of the firing rate of the population (see Figure 5-2C). To this end, we used the Fast Fourier Transformation on the mean SSCtx-LL firing rate (see Equation 5-4). This representation highlights low frequency oscillations in the SSCtx-LL firing rates which translate to waves of period 900 ms for WBM1 and 1666 ms for WBM2. These synchronized oscillations are also found in the mouse cortex in quiet states in vivo and in vitro [113]. During this state, the cortical neurons alternate in synchrony between periods of inactivity (UP state) and depolarizing periods (DOWN states). These oscillations might be the result of the balancing of activity between excitatory and inhibitory neurons. It has also been found that these oscillations might be related to the activity in other regions of the brain such as the Thalamus which is one of the major input contributors to the Cortex, as can also be seen in our model (see Sup. Figure 8A).

This simulation shows that the WBM in this configuration can replicate the activity of the SSCtx-LL in a quiet state.

### 5.3.2 Activity of the somatosensory cortex lower limb stimulated by the thalamus

In this section, we present the results of the simulation of the SSCtx-LL isolated from the rest of the brain for 4 seconds, stimulated by the ventral posteromedial nucleus of the thalamus (VPM) (see Section 5.2.2). These are shown in Figure 5-3. We observe in both networks an increase in spiking activity shortly following the stimulus of VPM. This activity consists, in both cases, of a first peak of activity 250 ms after the start of the stimulus. The activity will decrease to a stable state and remain in this state while the stimulus remains active. After the end of the stimulus, the firing rate activity returns to its resting state activity. 500 ms oscillations in firing rates of the excitatory population appeared in WBM1 once the network reached its resting state. In the case of the WBM1, the major change of activity was observed in the excitatory population which raised its firing rate to 10Hz, while in WBM2 this change was mostly suppressed by the inhibitory population. The inhibitory neuron population in the WBM2 takes a more important proportion of the neuron population than the WBM1 which explains this result.

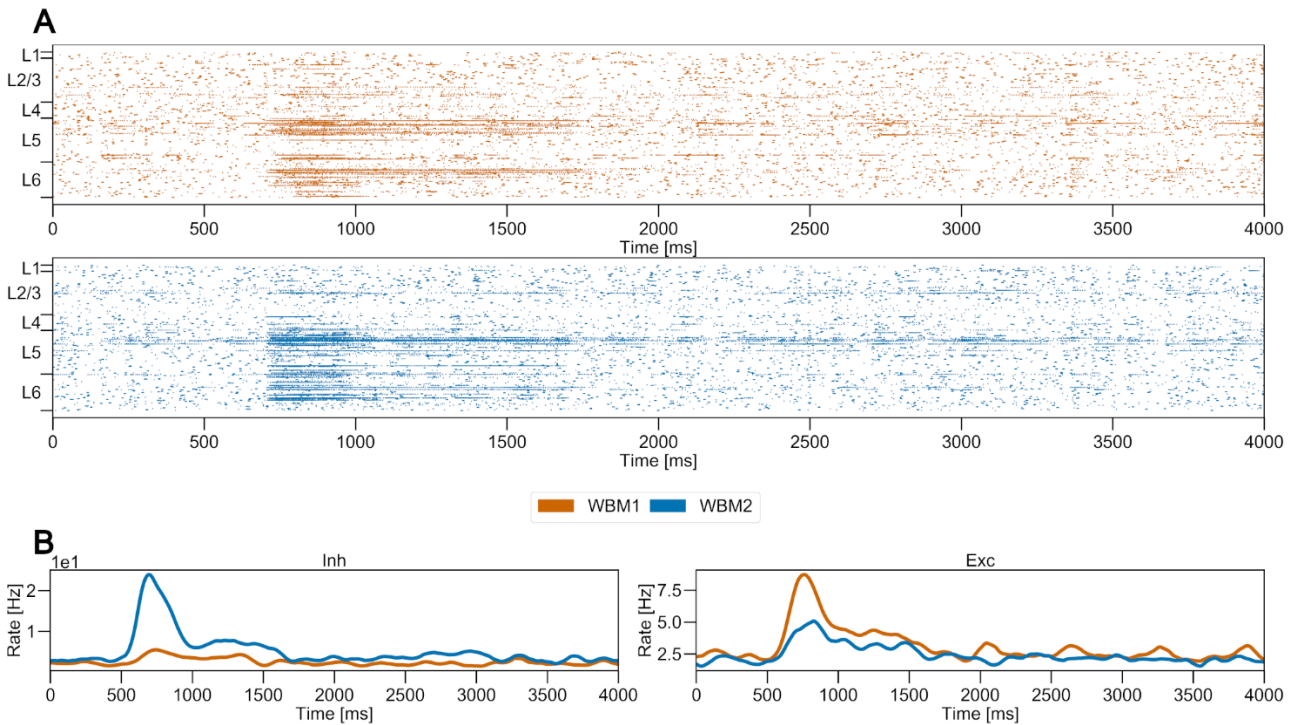


Figure 5-3: **Comparison of the spiking activity of the SSCtx-LL stimulated by the VPM between the WBM1 and WBM2.**

Comparison of the spiking activity of the somatosensory cortex lower limb (SSCtx-LL) between the two versions of the whole mouse brain model. In both cases, the network was stimulated by Poisson spikes mimicking the activity of the ventral posteromedial nucleus of the thalamus (VPM). Stimulation started at 500ms and ceased at 1500 ms. As before, WBM1 (in orange) corresponds to the version presented in Erö [2], while WBM2 (in blue) corresponds to our updated version of the network.

(A) Raster plot of the spiking activity in SSCtx-LL, sorted by layer. Neurons are distributed along the y-axis and each spike time is represented as a colored dot along the x-axis. The boundaries between layers are shown on the left.

(B) Mean instantaneous firing rate of the inhibitory and excitatory populations of the SSCtx-LL over time. The influence of the stimulation can be seen in both cases.

As in the resting state experiment, the network in this configuration produces spontaneous activity which reaches a stable state after a brief period of time. The perturbation introduced by the VPM is not sufficient to change its state. In the case of WBM2, the increase of the inhibitory population proportion, accentuated by a reduction of excitatory synapses coming from the VPM, have also significantly reduced the effect of the perturbation which means that the network is more resilient to noise. The response firing rate to the stimulus is also low compared to in vivo studies on rodents [4] but the delay of the response of the network seems to correspond to literature findings. These results might indicate that the weight of the connections between the VPM and the SSCtx-LL are too low or target the wrong population of neurons. We further investigated the density profile of the synapse counts from VPM to the SSCtx-LL (see Figure 5-4). Both in WBM1 and WBM2, we see that the projection primarily targets the layer 5 and layer 6 of the cortex which explain the strong reaction of these populations to the stimulus (see Figure 5-3A). However, these results do not perfectly match the density of bouton in the rat SSCtx-LL as measured by Meyer et al. [114]. The authors have shown that the Thalamus VPM targets

primarily the L3, L4 and to a lesser extent the L5b and L6. If the findings from Meyer et al. were to be confirmed in mice, we would have to integrate this connection probability rule in our connectivity recipe algorithm.

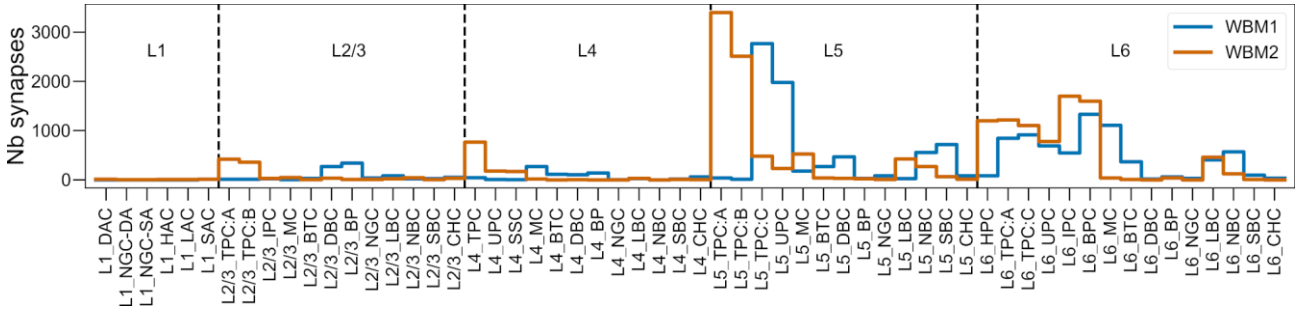


Figure 5-4: **Efferent synapses of the thalamus VPM targeting the SSCTX-LL, sorted by postsynaptic m-type.**

Distribution of the synapses between the thalamus VPM as source and the SSCTX-LL as target for each whole brain model version. The counts are sorted by layers and m-types.

### 5.3.3 Whole mouse brain activity

In this section, we present the results of the simulation of the whole mouse brain model for 4 seconds where each neuron was stimulated with a low frequency Poisson point process. The results of the simulation are shown in Figure 5-5 and Figure 5-6. We observe that the spiking activity of the different regions synchronize after the warmup of 500 ms period. The network in this configuration produces stable waves of activity of 800ms period for the WBM1 and 550ms for the WBM2. The spiking activity of each version seems to be mostly driven by the inhibitory neuron population (see Sup. Figure 7) whose firing rate is an order of magnitude larger than that of the excitatory population. This population in this network limits the activity of the excitatory population which can explain the oscillatory patterns. The oscillations of the network might also highlight, as pointed out by Erö [2], the role of adaptation to regulate the activity in the brain. Here, the short-term presynaptic dynamics and neuron adaptation of the network prevent self-sustained high activity in the network. In the case of WBM2, the activity is not entirely suppressed by adaptation and some regions remain highly active even in between wave peaks. This includes the cerebral nuclei (striatum and pallidum), the dorsal part of the midbrain, the frontal part of the isocortex, the ventral part of the thalamus (sensory-motor cortex related), etc. The change of proportion of inhibitory neurons in the network, due to the refinement of the cell atlas (see 0), plays a role here as the inhibitory neuron models are less affected by adaptation.

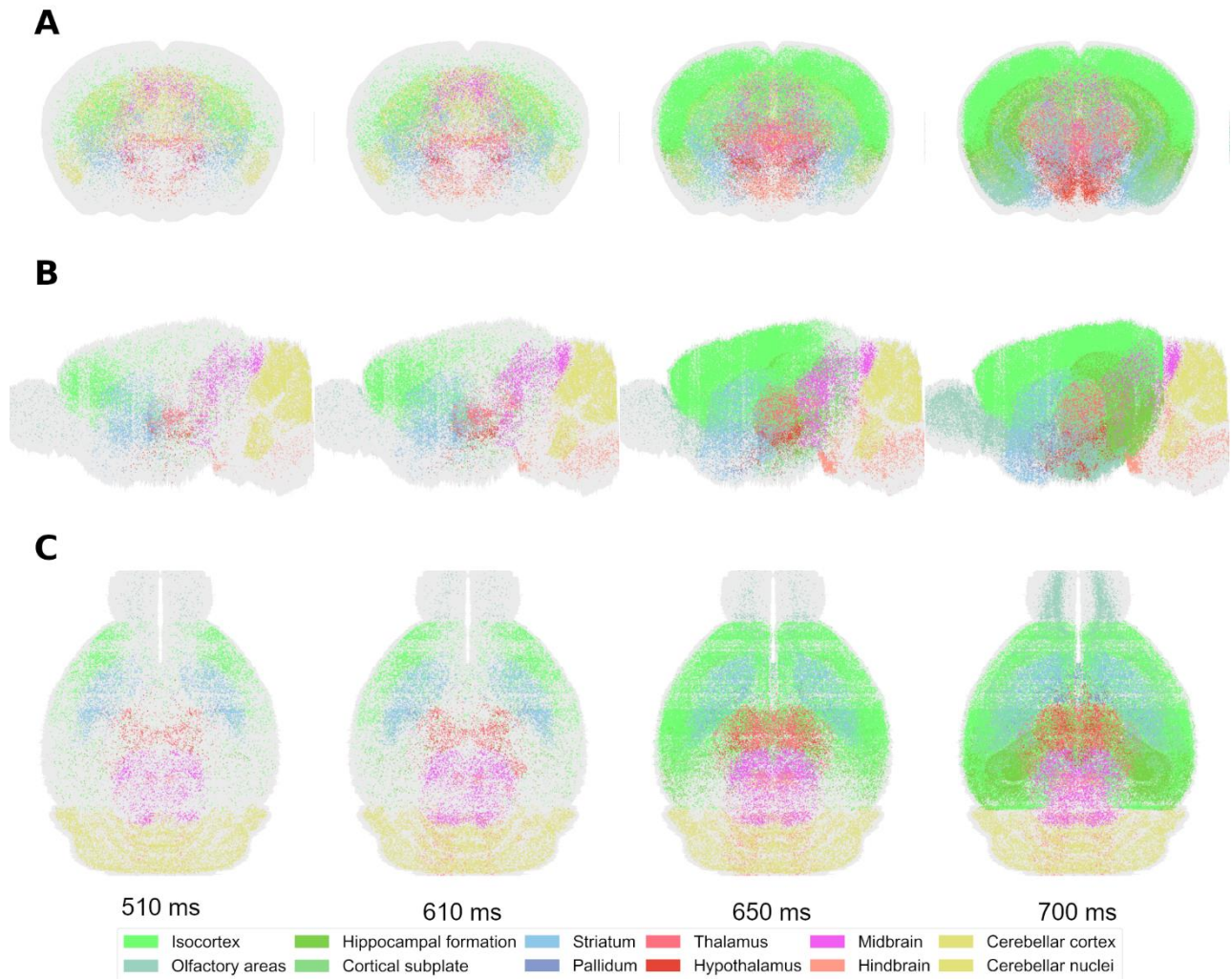


Figure 5-5: **Propagation of spiking activity in the whole mouse brain model.**

Spatial raster plots of the WBM2 spiking activity stimulated with a low frequency Poisson process, in coronal (A), sagittal (B) and axial (C) views at different points in time. Each neuron' spike is represented as a dot colorized according to the neuron brain region, as defined by the AIBS.

The change of activity in the olfactory areas is also most probably linked to the changes of the cell atlas as this region has seen its population of inhibitory neurons increase by more than 40%. In both versions, the activity of the cerebellum (cerebellar cortex and cerebellar nuclei) remains relatively constant throughout the simulation and despite the peaks of activity in the other regions. This region is quite isolated from the rest of the brain as is shown in Sup. Figure 10, which explains why its activity does not follow the drastic changes of the rest of the brain. On top of the changes of the cell atlas which increased the number of inhibitory neurons in the cerebellar cortex mostly, the change of the connectivity implemented in Chapter 4 within the cerebellum might play a role in the increase of firing rate observed in this region.

From Figure 5-5, we can also see how the wave of activity spreads in the WBM2. It is difficult to estimate from which brain region the wave activity takes its origin, but it seems to spread from the regions that remain highly active during the entire simulation (see also Sup. Figure 7D). Then, the activity propagates from the front of the brain towards its caudal regions. Finally, the wave activity reaches the hippocampal formation and the olfactory areas.



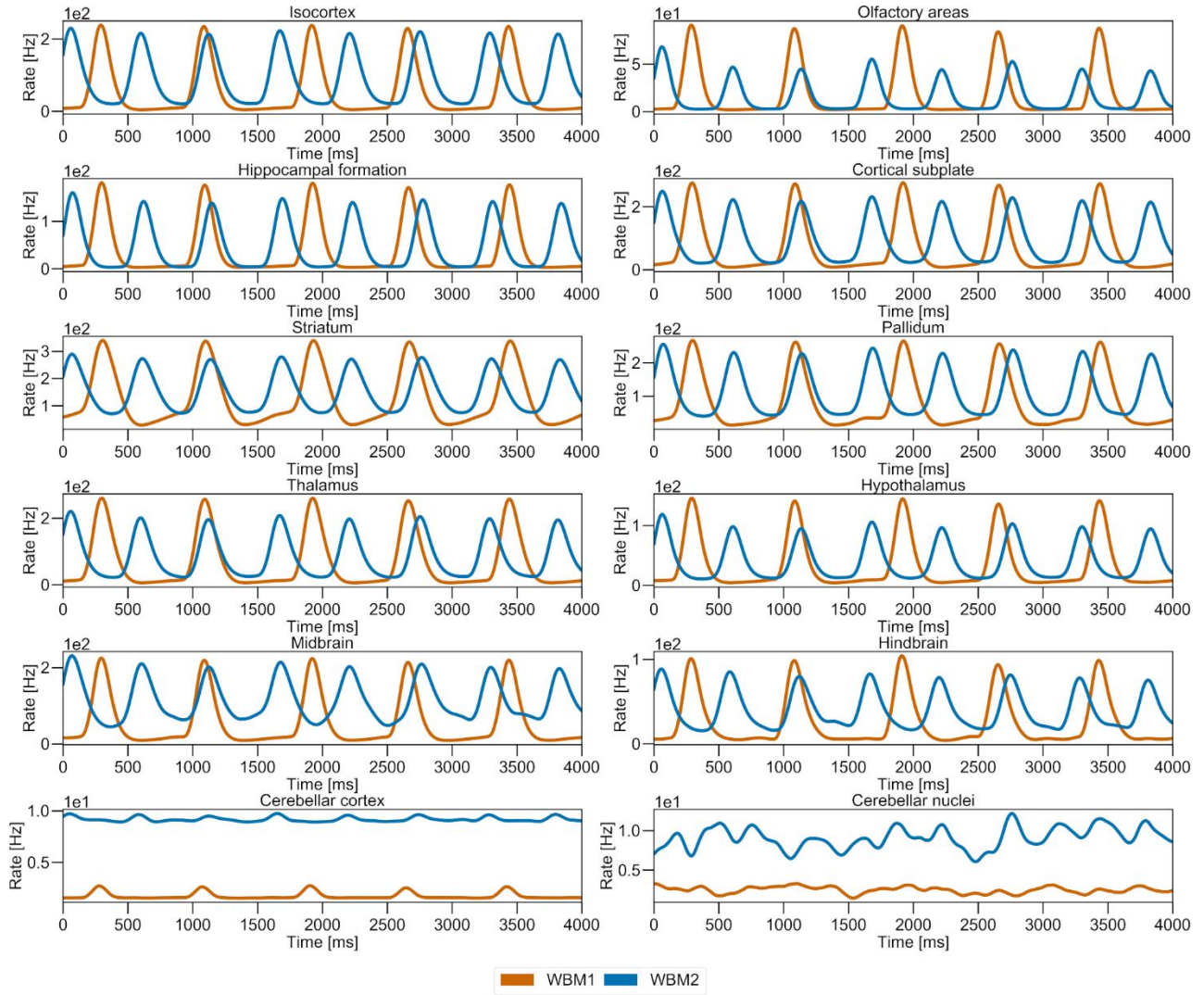


Figure 5-6: **Firing rate comparison for the different brain regions of the whole brain model.**

Mean instantaneous firing rate of the different brain regions of the whole mouse brain model when stimulated by a 0.5 Hz Poisson point process (see Section 5.3.3).

## 5.4 Discussion

### 5.4.1 Point-neuron network parameters

Our whole mouse brain point-neuron network (WBM) is based on our cell atlas and connectivity atlas described in respectively Chapter 3 and Chapter 4. The neuron type distributions are mostly based on the cell atlas and is refined in some regions with findings of literature. The electrical properties assigned to each neuron electrical type (e-type) is based on the database we presented in Chapter 4. To most of the e-types stored in our database were assigned the point neuron parameters from Naud et al. paper [104], which are chosen to produce the most common behavior of the neurons in the isocortex. Similarly for each synaptic type (Stype) assigned to a connection of our model is associated parameters from Markram et al. [4]. These parameters might be not adapted to certain neuron types or certain regions of the brain.

If detailed morphologic mouse brain tissue reconstructions are available in literature, we can use simplification methods, such as the Rössert et al.'s [100], to convert them into point-neuron networks which will capture the most of the dynamics of their original circuit. Another solution is to co-simulate the detailed circuit embedded into our scaffold whole brain point-neuron network. While this solution guarantees the best version of the whole mouse brain model, it will be more difficult to simulate efficiently as most neuronal circuit simulators are designed to simulate either detailed circuit with reconstructed morphologies (e.g., the NEURON simulator [115]) or point-neuron network (e.g., the NEST simulator [101]).

Different point-neuron models exhibit different dynamics in simulation, such as adaptation and dynamic or random spiking threshold. However, these dynamics require parameters which might not be available (e.g., adaptation parameters for the granule cells of the cerebellar cortex). On the other hand, certain properties from the literature might be difficult to integrate because they do not easily translate into point-neuron model's parameters. Future versions of the model might therefore replace the AdEx model assigned to certain neurons with a more advanced or simpler point neuron models to be able to integrate dynamics found in literature.

Our model implements short-term presynaptic dynamics from the Tsodyks-Markram model [105] which is sufficient to describe the activity of our connections in our simulations since our longest simulations lasted tens of seconds. If longer simulations of the models should be considered, long-term synaptic dynamics, such as long-term potentiation and long-term depression, might be required to replicate literature findings in simulation. These effects are triggered by a cascade of reactions at specific ion channels. They are therefore very specific and complex which might be difficult to capture as synaptic parameters in our model.

### 5.4.2 Simulations of the point-neuron mouse brain model

#### 5.4.2.1 *Simulation of the somatosensory cortex lower limb*

We simulated the somatosensory cortex lower limb (SSCtx-LL) of WBM first at resting state (see Section 5.3.1), and then stimulated by the ventral posteromedial nucleus of the thalamus (VPM - see Section 5.3.2). In both experiments, the SSCtx-LL was stimulated with a low background noise to mimic random spike input from the rest of the brain and spontaneous vesicle releases. We used the first experiment as a reference to scale the weight of the synapses of our model. This step was necessary as we simulated a scaled version of the point-neuron network (1% of the neurons and 1% of the synapses) which means that the parameters we assigned to the whole version might not be adapted to the scaled version. We decided to scale only the synaptic weight and determine its new value when the firing rate of each neuron population of the SSCtx-LL was below 10Hz to correspond to findings of Markram et al. [4]. For future simulations, we could however perform a more thorough exploration of different values to scale our point-neuron network parameters to see how they create different spiking regimes in simulation.

With this set of parameters, our model produced a stable spiking activity with low frequency waves of activity ( $\sim 1$ Hz). This state corresponds to the cortical activity observed on resting mice [102]. We notice, however, that the spiking activity of a portion of the L5 neuron population seems to be highly correlated with itself in both WBM1 and WBM2 (see Figure 5-1A). This might indicate that some neurons in this population are highly interconnected and therefore their spikes synchronize. When the SSCtx-LL network is then stimulated by the VPM, we observe that populations of the L5 and L6 layers see their activity raise (see Figure 5-3). These layers correspond to the targets of the VPM synapses. This increase of activity seems however to be restricted to the direct targets of the VPM (see Figure 5-4). These target neurons include inhibitory neurons which might quickly suppress the spreading of the stimulus to the other layer. It might also be that the scaling of the weights and number of connections was too strong to allow the transmission of spikes between neurons of the network. We also observe that the peak of spiking activity which was mostly due to the excitatory population in the WBM1 is now produced by the inhibitory population in the WBM2 (see Figure 5-3----B), indicating once again that the changes of the cell atlas (see Chapter 3) have a strong impact on the spiking activity of the resulting point-neuron network.

#### 5.4.2.2 *Simulation of the whole mouse brain*

In our whole brain simulation, we used the same configuration as for our resting state simulation of the SSCTX-LL, stimulating each neuron with a low background noise (see Section 5.3.3). We observe that the spiking activity synchronizes rapidly across the whole brain (see Figure 5-6). This creates an epileptic regime in the whole brain. Between WBM1 and WBM2, we observe that the size and period of the oscillations have been halved in almost all brain regions. We also see in the periods in between peaks of spiking activity, that WBM2 shows a higher firing rate compared to the WBM1. This means that a portion of the brain remains active continuously during the whole simulation. The Sup. Figure 7-AB shows clearly that the inhibitory population of the WBM2's minimum firing rate is around 20Hz and while the firing rate of the excitatory population, at the same time, is around 1Hz. Hence, a part of the inhibitory population of the WBM2 seems to be continuously active during the simulation, which might be the result of their stimulation by the background noise.

It is difficult to find the origin of a spread of spiking activity observed in the whole brain but with a spatial representation (see Figure 5-5), we can try to locate it. We can see that the cerebellar cortex remains active during the whole simulation. However, as we can see in Figure 5-6, the mean firing rate in this region is relatively constant during the simulation. The great number of neurons in this region might explain why this region seems to be constantly active in Figure 5-5. Next, the striatum is a region containing a great proportion of inhibitory neurons (see Figure 3-7 and Table 3-3) and might not be directly responsible for the increase in the spiking activity in the whole brain. As we pointed out previously, this region might be overstimulated by the background noise or the surrounding regions and participate to maintain a high firing rate even between peaks. On the other hand, the thalamus contains mostly excitatory neurons (see Figure 3-7 and Table 3-3). The region also sends axons to many regions of the brain (see Sup. Figure 10 and Figure 4-5), and it seems to remain active even in between peaks of firing rate (see Figure 5-5) due to the background noise. All these elements would indicate that high firing rates in the thalamus spread to the rest of the brain. Whatever the exact reasons for the oscillatory activity are, our anecdotal experiments clearly show that our method to stimulate the current whole-brain model is not yet accurate enough to reproduce the resting state activity of the whole mouse brain. Future experiments on the point-neuron model should focus on improving the stimulation of its neurons based on some experimental findings, such as resting state activity observed in brain regions.





# Chapter 6 Discussion

The goal of this work was to establish a data-driven pipeline to generate, simulate and validate point neuron networks of the whole mouse brain. Rather than targeting a particular property or behavior for our model, we want here to build a workflow that is based on clear assumptions and methods that allows the integration of a wide variety of data and literature findings into a coherent framework. We want our model to be used as a basis or *scaffold* to build more detailed circuit models or to provide a more realistic activity context to region models than the commonly used Poisson noise. The point neuron whole mouse brain model pipeline is a data-driven iterative workflow. It is meant to be regularly updated integrating more literature data, refining the assumptions made on the mouse brain and improving the methods used to build the model. The work presented here is the first major update to the model published by Erö [2] who has pioneered a first version of the pipeline to generate whole mouse brain models. In my thesis, I contributed several important improvements:

1. Methods to quantify and correct artifacts in the Nissl image stacks, yielding not only improved reference volumes, but also more accurate cell density estimations
2. A method to produce annotation volumes that combine the best features of the AV2 and AV3 from the AIBS.
3. Methods to determine the best set of Annotation and Nissl volume for the purpose of cell density estimation.
4. Methods to improve the orientation fields of brain regions in the brain computed from the AV.
5. Methods to integrate and consolidate a wide variety of literature data
6. All these improvements yield a greatly improved Cell Atlas that distinguishes many more inhibitory cell types than the first version.
7. Refinements to the micro- and mesoscale connectivity and a flexible database model for connectivity data.

## 6.1 Work achieved

### 6.1.1 Analyses and corrections of the reference atlases

The reference atlases consist of a pair of volumetric datasets (Nissl + Annotation volume) from the Allen Institute for Brain Science (AIBS). These datasets create a common coordinate framework (CCF) on which we build our point neuron whole brain model. We analyzed the different versions of the reference atlases released by the AIBS and created tools to detect and potentially correct the artifacts in these datasets.

For example, the Nissl volume from Dong [7] that is used in our pipeline to estimate cell densities shows artifacts such as holes, tears and over- or under-exposure. We described methods that mitigate most of these artifacts and that could be potentially applied to the isocortex. However, most correction methods also reduce local detail as they spatially smoothen the data. These tools are indeed based on the assumption of a certain homogeneity of Nissl expression within certain regions of the brain, but this assumption might not be valid in regions with large local differences in cell density, such as the Hippocampus CAs regions [43]. Future refinements of the Nissl volume will involve the integration of additional slice images of Nissl experiments available on the AIBS [21]. These images will provide an alternative reference on the intensity of certain slices of the Nissl volume, which can therefore be used to correct artifacts in the original Nissl volume. These images can be realigned to the Nissl volume using Krepl et al.'s pipeline [36].

The second reference atlas is an annotation volume (AV) that assigns a region to each voxel of the brain. This dataset is widely used by the neuroscience community to register experimental results in the mouse brain. However, multiple instances of these AVs exist to this date; each with its own advantages and disadvantages. In short, while older versions of the dataset (AV2, AV<sub>bbp</sub>) have more jagged edges along the rostral-caudal axis, the newest version (AV3) is smooth, but lacks important brain regions that were still present in previous versions. To obtain a version of the AV that is the best compromise, we re-aligned AV2 to AV3. Our algorithm uses ANTs tools [35] to realign the borders of each region of the brain individually, traversing the AIBS brain region hierarchy from the top-level regions to the leaf regions. This algorithm requires that reference and target AV have the same region hierarchy that is arranged in the same way. We therefore produced a merged version of AV2 and AV3 that uses the same set of region ids. However, some regions such as the hindbrain are so different between AV2 and AV3 that their realignment is still difficult for our algorithm.

Our future work on the annotation volumes will therefore involve the refinement of the merged AVs to simplify the task of realigning AV2 to AV3. We will also test smoothing algorithms such as the one proposed by Young et al. [45].

We also demonstrated that the alignment of the two reference atlases correlates with the quality of our estimates of cell densities in the mouse brain. Unfortunately, we noticed that this alignment is currently imperfect for each pair of reference atlas available. We proposed tools to compare the different versions and found that the best pair currently available for cell density estimation is the CCF<sub>bbp</sub> from Erö et al. [8]. The realignment of the Nissl volume to the Annotation volume is difficult as sometimes the boundaries between regions cannot be distinguished solely based on the Nissl volume. However, the boundaries between regions with a significant difference of cell densities can be easily determined on the Nissl volume (e.g., L1/L2 in isocortex and molecular/granular layers in the cerebellar cortex). We therefore plan to implement an algorithm to redraw the boundaries between these regions based on the Nissl volume.

We finally proposed an improved version of the orientation field algorithm developed by Erö [2]. This algorithm is used to orient cell morphologies or fibers in the brain towards a particular location, which is needed to reconstruct connectivity or to measure density profiles. Our new version of the algorithm corrects for the edge effects that were detected in the isocortex region. We used it to compute the orientation field in each region of the cerebellar cortex, and the isocortex. While our results corrected most of the issues detected in the previous version of the orientation field, the new version yields orientations fields that are locally less smooth, in particular when the AV has jagged contours. We will test the possibility of smoothing the orientation field in a future iteration of the pipeline. The orientation fields computed by our algorithm can also be used by our depth estimation algorithm that we presented to measure the distance of voxels in the brain towards a particular region of the brain. This was used to compute the depth of each voxel of the isocortex. However, this algorithm seems to perform less well on the edges of the regions. We will reuse our method to improve the orientation field computation to correct for these artifacts.

## 6.1.2 Refinement of the whole mouse brain cell atlas

Erö et al. [8] described a pipeline to estimate cell densities based on ISH datasets from the AIBS and a handful of literature values that constrain cell numbers in the whole mouse brain. This method yielded the first version of the blue brain cell atlas (BBCAv1). However, this approach predicts cell densities that are in some regions inconsistent with literature values. The authors argued that for most regions, there is no agreement in the literature and that values often vary by almost an order of magnitude. We conducted a comprehensive review of the literature on densities of inhibitory neurons in the mouse brain and observed indeed a great variability in cell counts, most probably due to differences in acquisition techniques.

We have therefore implemented a new method to estimate cell densities in the mouse brain, based on integrating and consolidating all available literature estimates available. We assumed that all estimates scatter around the ground truth and that the mean of all literature estimates should thus converge to this ground truth, the more values get integrated. We therefore scanned the literature to obtain as many estimates of inhibitory neuron counts as possible. For the other regions that were not covered by literature, we fitted transfer functions from ISH intensity to neuron densities to obtain density estimates in these regions. We then optimized the estimates obtained from literature and the fitting to make them coexist in the same cell atlas.

With this technique we could refine the inhibitory neuron composition of the Cell Atlas and provide estimates for new types of neurons, including PV+, SST+, VIP+ neurons, and remaining inhibitory neuron subtypes (BBCAv2). Additionally, the BBCAv2 can be easily extended to include more cell types. For some regions, the assumptions underlying our new pipeline can still lead to inconsistent cell density estimates (from literature or the transfer functions) that we have to correct with an optimization algorithm. Our future improvements will therefore target to reduce the errors that our analyses detected at the end of the pipeline. Future work will involve integrating more literature points as they become available, improving the input data from the AIBS, refining our assumptions locally, and finally considering new methods for the fitting.

The workflow that I have presented here is not restricted to inhibitory neuron types but can be applied to any pair of reference volumes and to any genetic marker dataset from the AIBS website, which means that this pipeline can be extended to other cell types, if enough cell density estimates from literature are available to allow a fitting of mean expression data.

This project aims to involve the scientific community to contribute with open access to data, software, and tools. Our results will be released at <http://bbp.epfl.ch/nexus/cell-atlas> (Blue Brain Cell Atlas website interface). The Blue Brain Cell Atlas version 2 is expected to be of use for many purposes. For instance, our model will allow experimentalists to understand regional composition and permit computational neuroscientists to place defined cell types in their simulations. Furthermore, it sets the stage for further subdividing

of inhibitory interneurons into more fine-grained subclasses and allows the neuroscience community to identify areas where current knowledge can be enhanced by additional constraints.

### 6.1.3 Refinement of the local connectivity in the whole mouse brain

The connectivity atlas provides a cell-to-cell connectivity matrix that connects the neurons of the cell atlas with synapses. The process of reconstructing the connectivity in the mouse brain is broken into two types of connectivity: the long-range and short-range connectivity. In this thesis, we refined the short-range connectivity to include literature findings, including connection types and their electrical properties. The number of brain regions and neuron types produce a large number of potential connections between any pair of neurons. We created a database to store cell types, connection types and their electrical parameters. We also created a Python interface for this database to be used in the process of generating the whole mouse brain model. We were also able to use this database to store the point neuron network parameters and replicate experiments from the literature.

The refined cell atlas also yielded a refined connectivity matrix. We analyzed this new connectivity matrix and implemented our new method to incorporate connection types from the literature. Our future work on the connectivity atlas will involve the integration of more whole-brain datasets focusing long-range connections in the brain such as the ones used in Zingg et al. [31]. We will also refine the cortical connectome by integrating the inter-region connectivity work of Reimann et al. [116]. Additionally, we would like to refine the short-range connectivity to take cell orientations into account, using the orientation field algorithm presented earlier.

### 6.1.4 Simulations of the whole mouse brain point neuron model

We were able to generate a new version of the whole mouse brain point neuron model (WBM2). Although a thorough analysis of the whole-brain model is beyond the scope of this thesis, we still performed a few benchmark simulations to compare the new model with the previous version of the whole-brain model (WBM1). The benchmarks involved 3 simulations of a scaled version of the point neuron models: a cortical resting state simulation, a simulation of the response of a cortical circuit to stimulus as well as a whole brain simulation with low background activity. We conducted a thorough analysis of the spiking activity of the WBM1 and WBM2. We have shown that the differences observed between the versions can be explained by the changes brought with the refinement of the cell atlas and the connectivity atlas. We managed to obtain a steady-state activity in the isocortex. We also managed to simulate the scaled whole brain model for four seconds with NEST. This simulation shows that some subregions in the brain produce a self-sustaining activity. This activity eventually spreads across the whole brain, activating almost every region before the adaptation and inhibitory neuron populations block it. This creates large oscillations of spiking activity in the whole mouse brain. With future versions of the WBM, we would like to refine the parameters of the stimulus assigned to each neuron of the brain to match local region findings. We also wish to analyze in more details the cluster of regions which were producing self-sustaining high activity during the whole brain simulations. These regions might receive too much stimulus from our background activity or be too tightly connected to each other.

## 6.2 Impact of our pipeline changes on the whole mouse brain model

In this thesis, my major contribution is a new method to estimate cell densities in the whole mouse brain based on literature findings and a clear list of assumptions (see Chapter 3). This refinement changes drastically our approach to estimate cell type densities. Instead of relying on a single value from the literature, from which we could estimate the full distribution of cell types in the brain as in Erö et al. [8], we are now integrating every local cell counts from the literature as constraints for our model. This new method facilitates the integration of new literature data and guarantees that the cell atlas model agrees with the literature in terms of cell type distribution in the brain. The constraints that we apply on the model to maintain its consistency also reduces the variability observed in literature (see Figure 3-10). On the other hand, it makes it more difficult to validate the final model as every literature value is used to constrain it. This difficulty to validate the final model is a clear drawback of our unbiased data-driven approach. We therefore must rely on qualitative comparison or statistical validation (see Section 3.3.5).

The changes to the cell atlas have a significant impact on the connectivity (see Chapter 4) and the simulation of the whole mouse brain point neuron model (see Chapter 5). The proportion of inhibitory neurons was indeed more important in almost every region of the brain (see Figure 3-7) when compared to our previous model. Now, our assumption is that the inhibitory neurons create local GABAergic connections. Overall, we therefore have more local inhibition in the new version of the whole brain model (see Table 4-1 and Table 4-3) which translate in simulation with a faster suppression of any strong activity in our whole brain neural network (see Figure 5-6). The waves of activity are therefore shorter in amplitude and period. From our whole brain simulation, we can also observe that the activity in the brain seem to spread between regions using mostly local connections (see Figure 5-5). While this spreading can be observed, for instance in regions of the cortex [117], some short inter-region connections might be unrealistic. For instance,

in our mouse brain model the striatum sends axons to the frontal part of the cortex due to the proximity of the regions. But, the striatum short range connectivity is mostly due to its interneuron population [95], which is not connected to the cortex. Our connectivity is therefore not specific enough with its targets. In our simulation results, we also see that the inhibition tends to suppress entirely any local activity for the entire population, resulting in an oscillating and epileptic regime in the brain. These oscillations quickly synchronize in almost every region of the brain, except for the cerebellum which spiking activity is stable over time and is very little affected by the other regions (see Figure 5-6). This can be explained by the poor coverage of the rAAV tracer injections targeting and originated from the cerebellum (see Figure 4-1). The waves of high neural activity seem to find their origin in regions of the brain that maintain a high activity during the entire simulation (see Figure 5-5). Among these regions, the thalamus is highly connected to the rest of the brain (see Figure 4-5 and Sup. Figure 10) and has a great proportion of excitatory neurons (i.e., approximately 80% of its neuron population according to our updated cell atlas), hence the rising part of the oscillating activity that we observe in the whole brain might be originated from there. The striatum maintains a high activity during the entire simulation which is related to its inhibitory population (which correspond to almost all its neural population) being constantly stimulated by its neighboring regions (isocortex and thalamus) but also by the background noise activity that we assigned to every neuron in the brain (see Section 5.3.3). In our future simulations of the whole brain, we should therefore further investigate the activity of the thalamus and striatum of the mouse brain in a resting state.

### 6.3 Assumptions of the whole brain model pipeline

Our model of the mouse brain is built upon a list of assumptions that we want to be clearly stated to be easily challenged and refined with new iterations of the model.

First our pipeline relies on a reference dataset pair (an annotation atlas and a Nissl volume). As we have extensively discussed in Chapter 2, the Nissl volume contains artifacts that are difficult to correct. It also represents a single mouse individual which might not be representative of the whole wild-type mice population. Improving the quality of Nissl volume using for instance some of the techniques discussed in Section 2.2.1 can play a role on the final estimates of our cell atlas. Still this dataset covers the entire mouse brain with a high precision and is aligned with the annotation atlas volumes. The annotations volumes (AV) from the Allen Institute for Brain Science (AIBS) are also not perfects. The precision of the region hierarchy is sometimes not sufficient to allow the integration of precise literature data (e.g., dorsal and ventral parts of hippocampal CA1). More some of the AV are very jagged along the rostro-caudal axis which makes it difficult to accurately measure the volume of each region. Finally, the resolution of these datasets is not sufficient to label correctly some regions of the brain (e.g., cerebellar cortex, Purkinje layer). We assumed for our pipeline that the alignment between the Nissl and annotation volume was good enough to estimate the cellular composition of the mouse brain. However, as we saw in Figure 2-5, this alignment is not perfect and has a significant impact on cell densities estimation. Future extensions of the model should therefore explore methods to improve the Nissl volume and related AV and their alignment with respect to one another.

Image stacks of *in-situ* hybridization (ISH) datasets from the AIBS were also used to estimate the density of cell types in the whole mouse brain. However, as we discussed in Chapter 3, these datasets do not cover the entire brain volume and some parts of the brain such as the olfactory bulb are sometimes left out. We should therefore try to improve the coverage of these datasets by combining multiple experiments or using more advanced interpolation techniques. Our cell atlas is built also upon assumptions about the cellular composition of the mouse brain. First, we assume that the Nissl and the ISH datasets from the Allen Institute show only the cell type of our interest. These datasets suffer however from artefacts due to acquisition methods. More, as we discussed in Section 3.4.4, this assumption does not hold for some specific regions and cell types. These exceptions can however be taken into account, changing our equations on the cellular composition for these regions.

Next, we performed a literature review for inhibitory neuron densities in the mouse brain. Some of the collected values from literature had to be transformed to be integrated in our model (e.g., transform cell counts into densities using the volume of the region). Each of these integrations therefore were based on the explicit assumption that the collected values could be fitted in our model and that the values used to transform them were correct. We collected values from the literature on wild type mice, yet we did not specify neither the specie model as it was done in Keller et al. [17], nor the sex, nor the method used to extract these cell counts. This is because we are trying to build a model of an average wild type adult mouse. It was also necessary to limit the constraints on literature values since we needed enough of them to build our model. In the future, we might consider to be more specific on our literature search especially if some acquisition techniques become more reliable.

As we refined the inhibitory neuron densities, we assumed the cell and neuron density estimates obtained with the method from Erö et al. [8] to be correct. However, these values were obtained using from global values from literature which means that the model

might be in contradiction with literature locally. We could refine these steps of the pipeline, using our new method to estimate the inhibitory neurons densities.

We assumed the me-type distributions from Markram et al. [4] could be applied to the entire isocortex of our model to refine the cell type distribution from our cell atlas. We also integrated the me-type to me-type connectivity matrix from the same paper in our connectivity atlas. The cellular distributions and connectivity matrix from Markram et al. come however from experiments on the rat somatosensory cortex lower limb which might not be adapted to the entire isocortex of the mouse brain. To each neuron without an e-type defined by a literature source, we assigned a default e-type from Naud et al. [104]. Similarly, we assigned a default s-type to each connection types for which we could not find literature information. These assumptions should however be challenged to refine the parameters of the resulting point-neuron whole mouse brain model (WBM).

We simulated a scaled version of the WBM on NEST. We decided to scale the network by reducing the number of neurons and synapses, as suggested by Erö [2], but also the scaling of the synaptic weight. This new synaptic weight value was chosen based on the results of simulation in the somatosensory cortex lower limb, which might not be correct for other regions. We used Poisson processes with a low frequency to stimulate each neuron of our circuits during our simulations. This was our solution to mimic the background spiking activity in the brain and obtain a resting state activity. However, the parameters of this stimulus should be discussed as they might not be adapted in every region of the mouse brain or in certain simulation configurations.

In our pipeline we tried to clearly state and be aware of all the assumptions we take to build our model. Each of them should be challenged at each refinement of the pipeline in order to reduce the potential errors that they might introduce.

## 6.4 Challenges related to a data-driven integration of literature data

In this thesis we proposed a new data-driven method to integrate the literature data into a model of the mouse brain. In opposition with most of the approaches to reconstruct the whole mouse brain, our technique does not select literature value to constrain the model but rather leverage all values available in the literature. Literature on the cellular composition of the brain is indeed very variable due mostly to differences in the methods to count cells in the brain. The wide variety of techniques used to estimate cell counts makes it very difficult to evaluate the quality of the values reported in literature, even for the authors of each experiment. More, some studies are old and the original data or the information on the methods used might be lost. Our solution is therefore to average of all similar literature values without giving a weight to the different methods used to obtain them. This way we avoid a bias of representation of a particular method or authors. We can nonetheless analyze the literature values gathered to determine if the improvement of the acquisition techniques or the usage of a better annotation atlas produces a systematic drift of estimation to correct some of the other techniques. We assume that the literature estimates are sampled around the ground truth, which means that our model should converge with the integration of more and more literature values. However, doing so, we take the risk of sampling the variability of the literature values extracted. Our results on the cell atlas show that we managed to reduce the variability of the results by integrating more literature values in the model. This is due to the constraints that we impose on our cell estimates, between regions and cell types. On one hand, these constraints limit the field of possibility of our cell count estimates reducing the variability of our final model. On the other hand, the constraints might strongly override literature estimates to make them fit into the model. Sometimes this overriding of literature values might drastically change them. These drastic changes are easy to spot and can therefore trigger more precise analysis on the composition of certain regions of the brain, to assert that our assumptions make sense in this context.

My efforts to improve the data-driven pipeline have generated a new version of the point-neuron model of the whole mouse brain. With this iteration, I have provided more analysis of the datasets consumed by the pipeline, simplified the integration of literature data and provided new leads to explore to further extend and refine the model. The pipeline and its tools are expected to be widely used by the scientific community. They provide a reference framework in which neuroscientists could integrate their own reconstructions of the mouse brain, identify areas where current knowledge can be enhanced by additional constraints and eventually collectively converge on a detailed data-driven model of the mouse brain capable of reproducing experiments performed on real mice.



# Supplementary materials

## Review of literature on densities of inhibitory neuron in the mouse brain

We performed a systematic review of the literature for inhibitory neurons in the mouse brain. To do so, we mostly used the google scholar tool to find articles. Our searches involved combinations of the following key words (and their respective common abbreviations, e.g.: GAD67 for GABAergic): mouse, neurons, cells, inhibitory, GABAergic, parvalbumin, somatostatin, vasoactive intestinal peptide, interneurons, quantitative analysis, counts, densities, etc. We additionally searched more specifically for specific regions of the brain to confirm findings or try to improve our coverage of the whole brain. Except for the striatum, we only selected papers on mice experiments. We also made sure to select a single paper per experiment to prevent any duplicate referencing. We extracted the mean value provided in each paper and its standard deviation whenever this value was available, Standard Error of the Mean were converted to standard deviations based on the numbers of individuals used to make the estimates. A density estimate in a specific brain region was assumed to be valid for the entire region. If the estimates were provided in the form of counts, we converted these into densities using the region volume from the paper, if available, or computed from the annotation volume (AV). When ratios (or percentages) of neuron type in a region were provided as proportions of the local cells or neurons populations, these ratios were multiplied by the corresponding densities from the BBcAv1 [8]. In total, we extracted density values based on 43 different papers [8], [13], [14], [40], [42], [43], [50]–[52], [56], [58], [74]–[77], [83], [84], [88], [90], [93]–[95], [118]–[138].

For each paper providing regional counts, densities, or proportion of inhibitory neuron types, we assigned the closest matching AV region(s). Sometimes several regions were assigned to the same literature value, while sometimes multiple values from literature were averaged to fit into a region of the AV. For instance, the cortical layer 6 is subdivided into two sublayers a and b in the AV but some papers do not make this distinction [42], [126]. For these cases, we assumed the same densities for both subregions.

Several papers provided densities for the frontal cortex [42], [90], [123]. This region was assumed to cover the prefrontal area, the frontal pole and infralimbic area of the AV.

We extracted density values from the Table 2 of Suzuki and Bekkers [94] in the anterior part of the piriform area of the olfactory areas (i.e., the part in contact with the lateral olfactory tracts). In absence of a better We assumed these estimates to be true in the entire piriform area. The layer 1, 2 and 3 were respectively associated to the molecular layer, the pyramidal layer and the polymorph layer. The a and b subdivisions of the layers 1 and 2 were not present in the AV so we averaged the two estimates for each layer.

Xu et al. [42] reported densities in the layer 4 of the frontal cortex. But this layer does not appear in the frontal part of the cortex of the AV, which is why the value was ignored.

The globus pallidus densities in Gourfinkel-An et al. [40] were assumed to be the same in both its internal and external segment subdivisions in the AV. A global estimate of inhibitory neurons in the entire cerebellar cortex was estimated from Gourfinkel-An et al. based on the authors estimates in the molecular and granular layers. These estimates were multiplied by the proportion of each layer volume in the cerebellar cortex. No voxels of the AV are labelled as belonging to the Purkinje Layer so its contribution to the cerebellar cortex density estimate was null.

Irintchev et al. [138] reported inhibitory neuron densities which we extracted from their Figure 3. However, it was not very clear from which region of the somatosensory cortex one of their estimates were taken from. Hence, we considered these estimates for the entire region.

From Zhao et al. [120] Figure 3 we manually extracted the counts and standard deviation of GAD67 positive cells in coronal slices of the lateral septal nucleus. The subregions from the paper are more precise than the AV. We averaged the intermediate counts values with the rostro-ventral counts and assigned them to the rostroventral region of the AV. The caudal-ventral counts from the paper were assigned to the ventral part of the AV as the voxels assigned to this region are located at the most caudal part of the lateral septal nucleus. The values left of the paper were averaged and assigned to the caudal-dorsal part of the AV. Counts were then converted to densities based on the volume of the counting frame (i.e.,  $163 * 163 * 30 \mu\text{m}$ ).

---

Han et al. [77] provided counts of GAD67 neurons from whole brain 30  $\mu\text{m}$ -thick coronal slices. The mean surface occupied by the regions of the article in the coronal slices were estimated from similar coronal slices in the AV. The resulting densities are very low when compared to other literature sources. These values were therefore not considered in the fitting process.

Ono et al. [129] provided densities of GABAergic neurons in cells / 104  $\mu\text{m}^2$  from 50  $\mu\text{m}$ -thick transverse slices of the mouse brain. No standard deviation value was provided in the paper.

The striatum GABAergic neuron densities were estimated from a collection of papers on the rodent [80], [139]–[143], see review from Tepper et al. [95]. From these papers, we could estimate the total proportion of inhibitory neuron in the striatum as the sum of the proportions of each of its distinct GABAergic cell type:

- Medium spiny cells: 95% [80].
- Cholinergic: 1.7% [139].
- Parvalbumin: 0.7% [140], [141].
- Calretinin: 0.8% [140].
- Neuropeptide Y: 0.9% [142].
- Tyrosine hydroxylase: 0.2% [143]

The total percent of inhibitory neuron in the striatum is therefore at least equal to 99.3%, which is the value we used in our pipeline.

In Gotts et al. [122], GAD67 immunoreactive neurons were counted in 50  $\mu\text{m}$  sections of the nucleus of the solitary tract of the mouse brain (see their Table 1). We converted these counts into densities, using the mean volume of 50  $\mu\text{m}$  coronal slices of the nucleus of the solitary tract in the AV. The paper's estimates in the subregions of the nucleus of the solitary tract were matched with their closest region in the AV:

- Medial part (paper)  $\rightarrow$  Gelatinous part (AV)
- Central part  $\rightarrow$  Central part
- Commissural part  $\rightarrow$  Commissural part
- Dorsomedial, Intermediate, Medial parts  $\rightarrow$  Medial part
- Interstitial, Lateral, Ventrolateral, Ventral parts  $\rightarrow$  Lateral part

We took the sum of the left and right hemisphere as this distinction does not exist in the AV. We also averaged all values assigned to the lateral and medial subregions in the AV.

Counts of GAD67 neurons in the mouse nucleus of the solitary tract from 10 sections (100 \* 100 \* 50  $\mu\text{m}$ ) were manually extracted from the Figure 1F in Wang and Bradley [135].

In Okada et al. [136], densities of GAD67 neurons in the mouse nucleus of the solitary tract were measured by the authors in 4 areas of the regions. The average of the densities in the areas was taken, as the areas taken by the authors did not correspond to any of the AV subregions.

Lefort et al. [41] provided counts of inhibitory neurons in a cylinder of the C2 barrel column of the mouse isocortex. We transformed the counts into densities based on the dimensions of the cylinder provided by the authors. We additionally applied the densities in the C2 barrel column to the whole barrel region (including the septa).

Almási et al. [126] provided inhibitory neuron densities in the somatosensory cortex, barrel field. The estimates from the layer 5a and 5b in the paper were averaged as the AV does not have this layer subdivision.

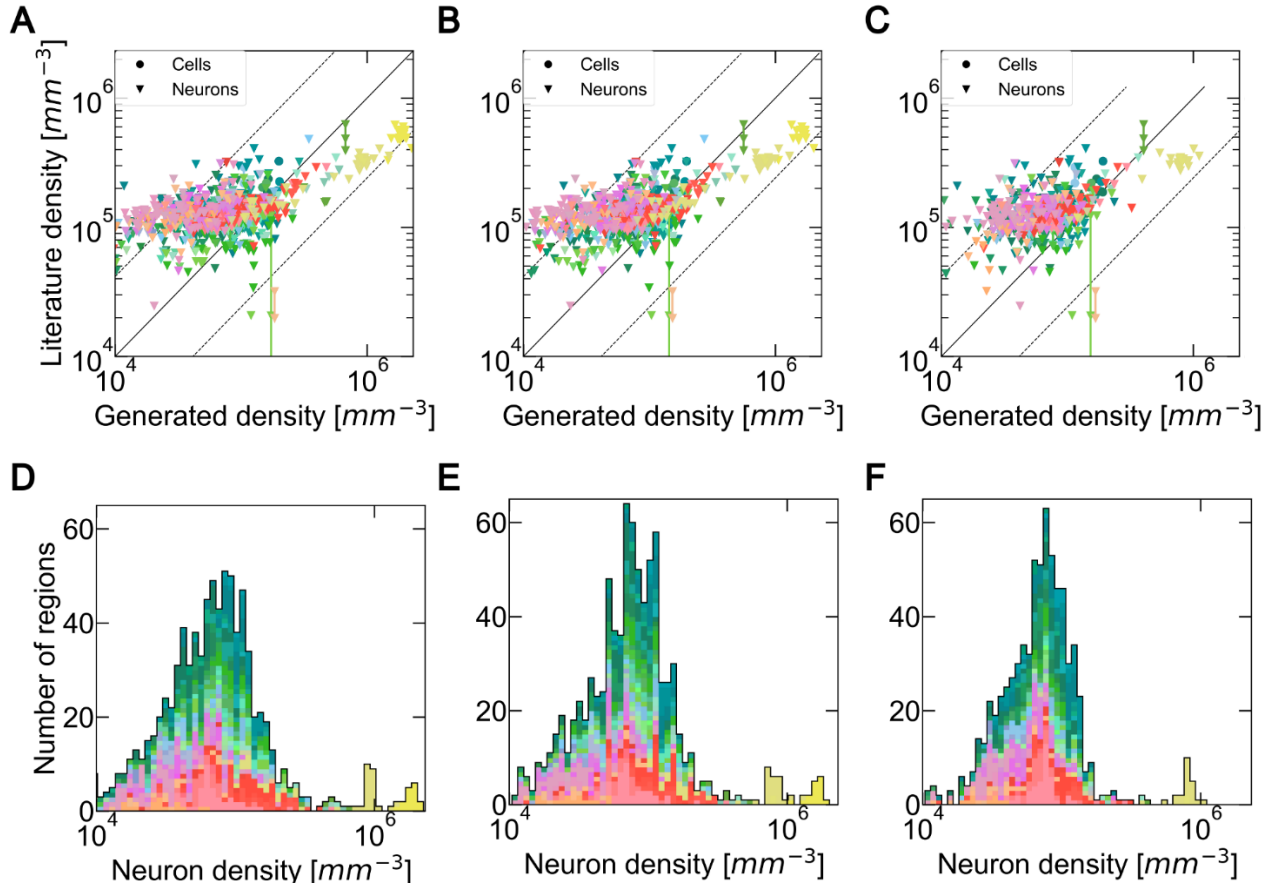
Jinno and Kosaka [119] and Whissell et al. [123] reported densities in the dorsal and ventral parts of the hippocampus. This subdivision of the hippocampus is not present in the AV, and as a result, the values reports were averaged to represent the full region. Additionally, estimates from the lateral and medial subdivisions of the visual areas from Whissell et al. were averaged for similar reasons.

Hafner et al. [121] provided counts of VIP+ neurons in the somatosensory cortex, barrel field. These counts were converted into densities using the size of the area used to count the cells (0.1 mm \* 0.24 mm) multiplied by the mean depth of the region, based on measurements from Lefort et al. [41].

From Erö et al. [8], we finally extracted the neuron densities in regions reported in literature to be fully inhibitory. These includes the layer 1 of the isocortex [13], the molecular layer of the cerebellar cortex [14], the reticular nucleus of the thalamus [133].



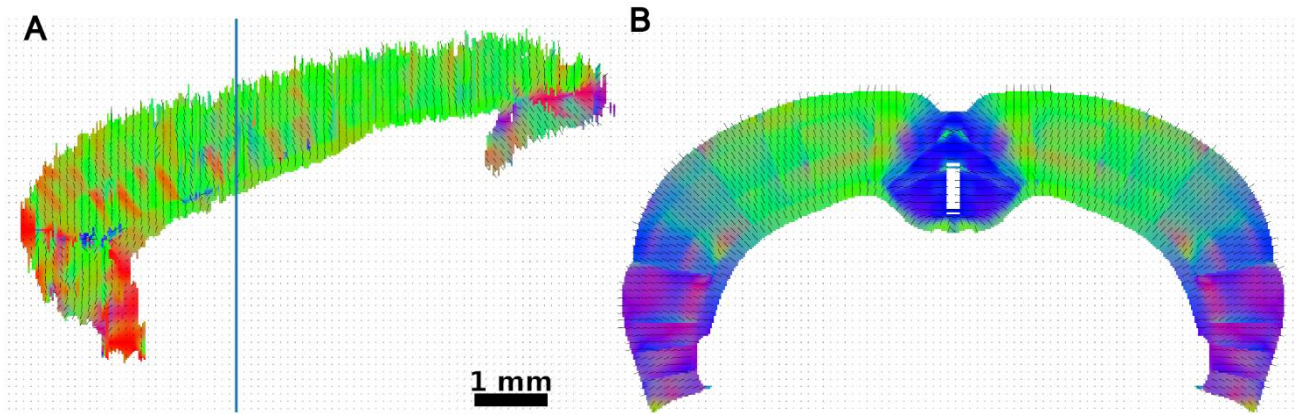
## Supplementary figures



Sup. Figure 1: **Global impact of the annotation atlas on cells and neuron densities.**

(A), (B), (C) Distribution of cell and neuron density values reported in literature against generated densities for similar regions, using the 3 pairs of annotations and Nissl volumes (from left to right: Erö et al. 2018, CCFv2 from the AIBS, CCFv3 from the AIBS). When multiple literature sources are available for the exact same region, they are both shown as a data point and are linked together. The color encodes the brain regions according to the AIBS region color palette, while the shapes of the points encode for cell types. The middle line delimits equal quantities, while the dashed line shows the average deviation of 2.7-fold between literature values reporting on the same region. Some subregions of the brain are not represented in CCFv3 which explains the different numbers of points.

(D), (E), (F) Histogram of the brain regions in terms of neuron density values for the 3 pairs of annotation and Nissl volumes (from left to right: Erö et al. 2018, CCFv2 from the AIBS, CCFv3 from the AIBS). Each region is represented with a single-color patch of the same size.



Sup. Figure 2: **Cortical orientation field computed from each cortical subregion individually.**

Sagittal view and coronal views of the isocortex orientation field. To each voxel of the isocortex in the  $AV_{BBP}$  is assigned a 3D vector defined with the method described in Section 2.2.4, but treating each subregion of isocortex individually. The Red Green and Blue proportions of the voxel colors correspond to the orientation vectors norm on respectively the coronal, axial and sagittal plane, black lines are their projected axis. Colors represent the orientation vectors norm on their respective plane, black lines their projected axis. The blue line in (A) shows the position of the coronal slice shown in (B).

We ordered  $\mathbf{R}$  according to the regions depth in the region hierarchy of the AV to obtain  $\mathbf{R}_O$ .  
Confidence interval of  $nPV$  (similarly for  $nSST$ ,  $nVIP$ , and  $nGAD$ ),  $\forall r \in \mathbf{R}$ :

$$[minPV_r, maxPV_r] \text{ where } \begin{cases} minPV_r = nPV_r - stdPV_r \\ maxPV_r = nPV_r + stdPV_r \end{cases}$$

```

1 foreach  $r \in \mathbf{R}_O$  do
2   if  $nNeu_r < nGAD_r$  then
3     |  $nGAD_r = nNeu_r$ 
4   end
5    $maxGAD_r = \min(maxGAD_r, nNeu_r)$ 
6    $minGAD_r = \max(minGAD_r, nNeu_r)$ 
7    $sumPSV = nPV_r + nSST_r + nVIP_r$ 
8   if  $nNeu_r < minPV_r + minSST_r + minVIP_r$  then
9     /* case where no solution can be found within confidence intervals,
10      PV, SST and VIP proportions are kept
11      */
12      $nPV_r = minPV_r = maxPV_r = nNeu_r \times nPV_r / sumPSV_r$ 
13      $nSST_r = minSST_r = maxSST_r = nNeu_r \times nSST_r / sumPSV_r$ 
14      $nVIP_r = minVIP_r = maxVIP_r = nNeu_r \times nVIP_r / sumPSV_r$ 
15   end
16   else if  $nNeu_r < sumPSV$  then
17     /* case where there is a solution within confidence intervals,
18      but mean estimates are greater than the number of neuron
19      */
20      $remainingSum = sumPSV$ 
21     while  $nNeu_r < sumPSV$  do
22        $ppv = nPV_r / remainingSum$ 
23        $psst = nSST_r / remainingSum$ 
24        $pvip = nVIP_r / remainingSum$ 
25        $diff = sumPSV - nNeu_r$ 
26        $nPV_r = \max(minPV_r, nPV_r - ppv \times diff)$ 
27        $nSST_r = \max(minSST_r, nSST_r - psst \times diff)$ 
28        $nVIP_r = \max(minVIP_r, nVIP_r - pvip \times diff)$ 
29        $remainingSum = sumPSV = nPV_r + nSST_r + nVIP_r$ 
30       if  $nPV_r = minPV_r$  then
31         |  $remainingSum = remainingSum - nPV_r$ 
32       end
33       if  $nSST_r = minSST_r$  then
34         |  $remainingSum = remainingSum - nSST_r$ 
35       end
36       if  $nVIP_r = minVIP_r$  then
37         |  $remainingSum = remainingSum - nVIP_r$ 
38       end
39     end
40      $maxPV_r = nPV_r$ 
41      $maxSST_r = nSST_r$ 
42      $maxVIP_r = nVIP_r$ 
43   end
44   else
45     /* case where the mean estimates are correct,
46      only maximums have to be checked.
47      */
48      $maxPV_r = \min(maxPV_r, nNeu_r - nSST_r - nVIP_r)$ 
49      $maxSST_r = \min(maxSST_r, nNeu_r - nPV_r - nVIP_r)$ 
50      $maxVIP_r = \min(maxVIP_r, nNeu_r - nPV_r - nSST_r)$ 
51   end
52 end

```

Sup. Figure 3: **Algorithm 1: Cap inhibitory densities to number of neurons.**

The estimated counts of GAD67+ neurons and the sum of PV+, SST+, VIP+ neurons counts are limited by the previously computed neuron counts (step 2 of the BBCAv2 pipeline - see Figure 3-1) to ensure that assumptions 2 and 3 are fulfilled (see Section 3.2.4). Recall that  $\mathbf{R}$  is the set of brain regions,  $\forall r \in \mathbf{R}$  inversely ordered according to their depth in the region hierarchy of the AV ( $\mathbf{R}_O$ ), the algorithm checks if the conditions  $nNeu_r \geq nGAD_r$  and  $nNeu_r \geq nPV_r + nSST_r + nVIP_r$  are satisfied. If not, it finds a solution which tries to match the following properties, ordered by priority: (1): Remain in range of the standard deviation of each value (confidence intervals), (2): Maintain the proportion of PV, SST and VIP within the region. If a solution exists within the confidence intervals, then the number of extra neurons (diff variable at line 19) is subtracted proportionally to the ratio of each neuron type (computed at lines 16-18). Then, if one of the neuron type estimates reaches the minimum of its confidence interval, then it is no more reduced (lines 20-22) and the remaining extra neurons are subtracted from the other neuron type estimates (lines 25, 28, 31).

We ordered  $\mathbf{R}$  according to the regions depth in the region hierarchy of the AV to obtain  $\mathbf{R}_O$ .  
Confidence interval of  $nPV$  (similarly for  $nSST$ ,  $nVIP$ , and  $nGAD$ ),  $\forall r \in \mathbf{R}$ :

$$[minPV_r, maxPV_r] \text{ where } \begin{cases} minPV_r = nPV_r - stdPV_r \\ maxPV_r = nPV_r + stdPV_r \end{cases}$$

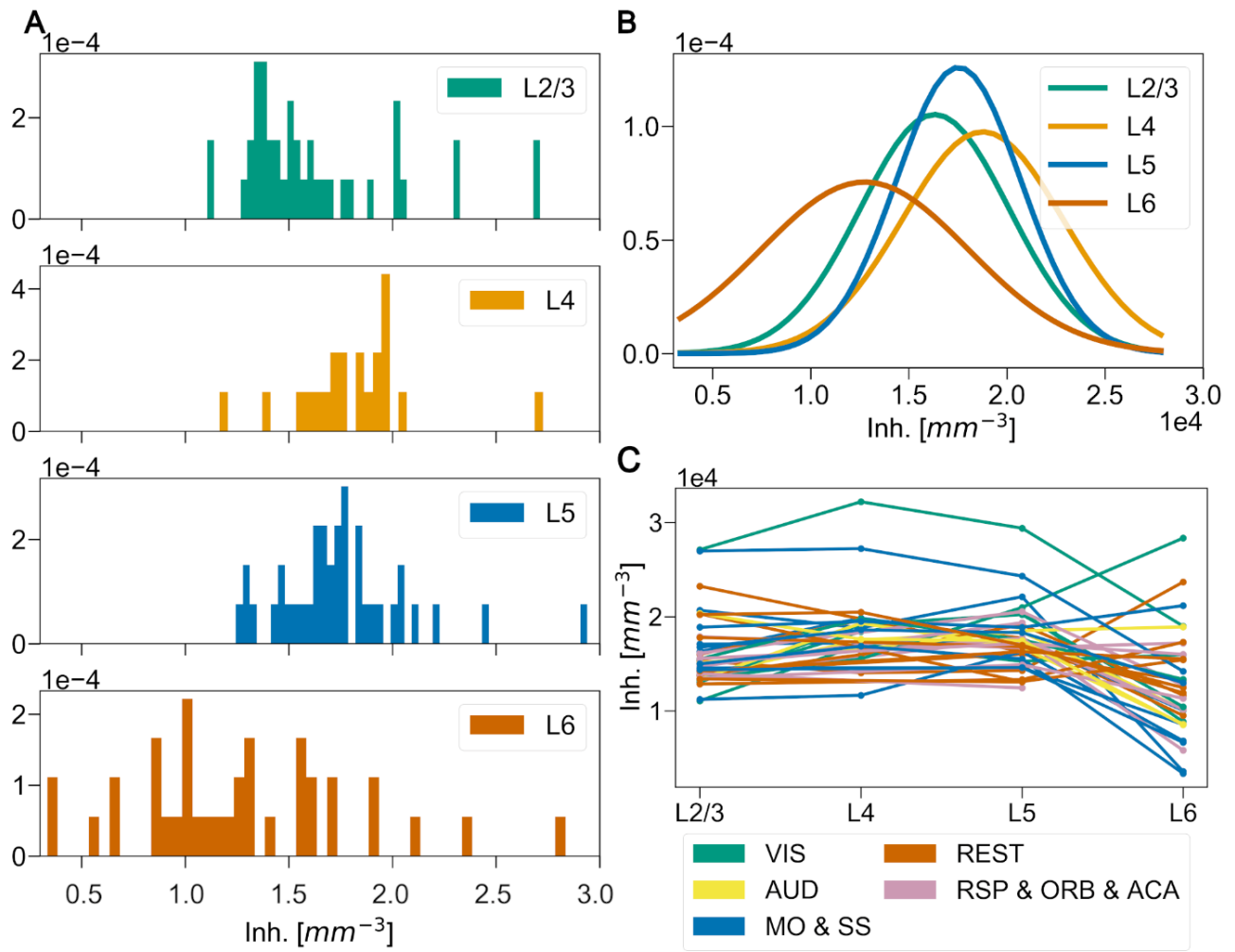
```

1  foreach  $r \in \mathbf{R}_O$  do
2       $sumPSV_r = nPV_r + nSST_r + nVIP_r$ 
3      if  $sumPSV_r > nGAD_r$  then
4           $q = (sumPSV_r - nGAD_r) / (sumPSV_r - nGAD_r + maxGAD_r - minSumPSV_r)$ 
5          // If  $q \leq 1$  there is a solution within confidence intervals
6           $nGAD_r = nGAD_r + q \times stdGAD_r$ 
7           $nPV_r = nGAD_r \times nPV_r / sumPSV_r$ 
8           $nSST_r = nGAD_r \times nSST_r / sumPSV_r$ 
9           $nVIP_r = nGAD_r \times nVIP_r / sumPSV_r$ 
10         end
11          $maxGAD_r = \min(maxGAD_r, nGAD_r)$ 
12          $minGAD_r = \max(minGAD_r, nGAD_r)$ 
13          $maxPV_r = \min(maxPV_r, nPV_r)$ 
14          $minPV_r = \max(minPV_r, nPV_r)$ 
15          $maxSST_r = \min(maxSST_r, nSST_r)$ 
16          $minSST_r = \max(minSST_r, nSST_r)$ 
17          $maxVIP_r = \min(maxVIP_r, nVIP_r)$ 
18          $minVIP_r = \max(minVIP_r, nVIP_r)$ 
19     end

```

Sup. Figure 4: **Algorithm 2: Maintain inhibitory densities coherence.**

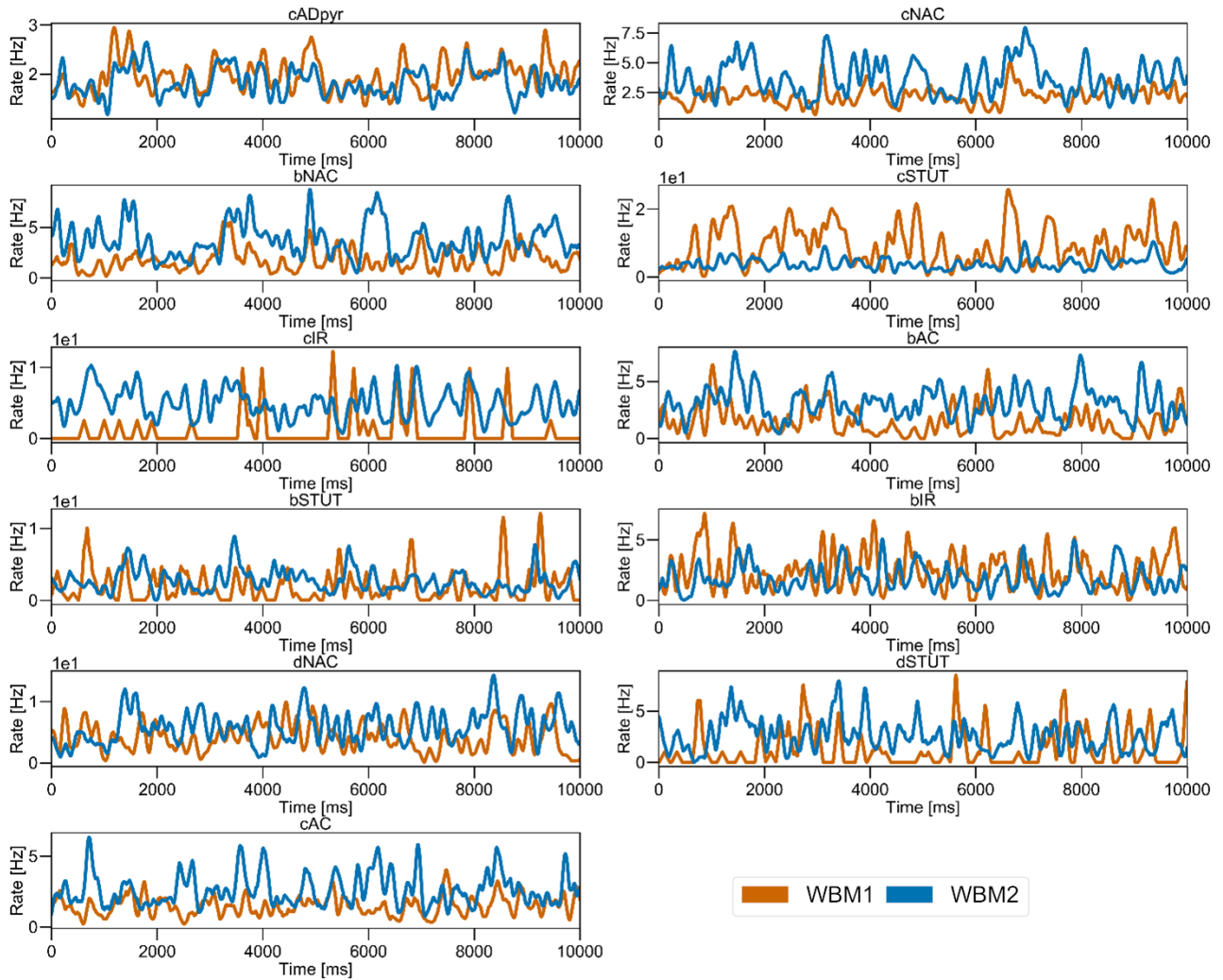
This algorithm corrects the estimated densities of PV+, SST+, VIP+ and GAD67+ neurons in the model so that assumption 4 of Section 3.2.4 is fulfilled.  $\mathbf{R}$  is the set of brain regions,  $\forall r \in \mathbf{R}$  inversely ordered according to their depth in the region hierarchy of the AV ( $\mathbf{R}_O$ ), the algorithm checks if the condition  $nGAD_r \geq nPV_r + nSST_r + nVIP_r$  is satisfied. If not, it finds a solution which tries to match the following properties, ordered by priority: (1): Remain in range of the standard deviation of each value (confidence intervals), (2): Maintain the proportion (ratios) of PV, SST and VIP. When a set of value exists so that the property (1) is fulfilled, there is a correction factor  $q \in [0,1]$  which corresponds to the fraction of standard deviation needed to guarantee that the sum of the inhibitory subtypes remains under the estimated count of inhibitory neurons and that each value remains within its confidence interval.



Sup. Figure 5: **Cortical inhibitory density distribution.**

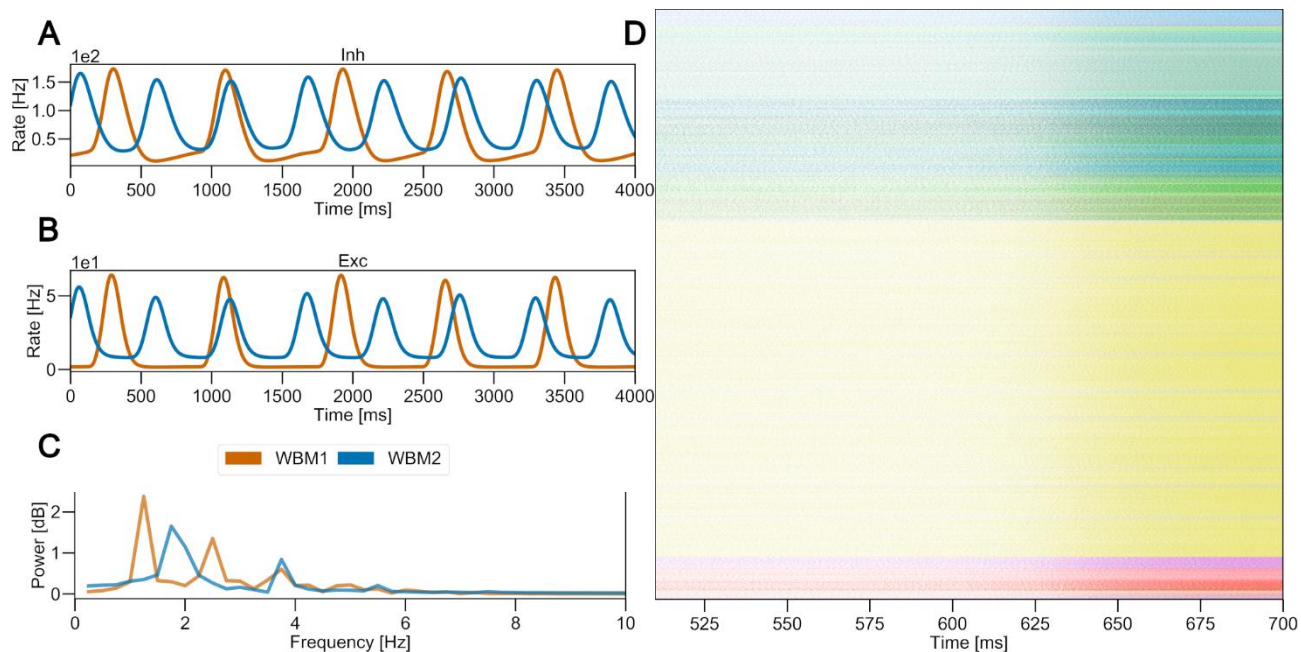
(A) Normalized distribution of cortical inhibitory neuron densities across the isocortex subregions grouped by layers. For each layer, the fitted normal distribution associated is displayed as a line on top of it. The normal distributions of L2/3 (mean  $1.6 \times 10^4$ , std.  $3.8 \times 10^3$ ), L4 (mean  $1.9 \times 10^4$ , std.  $4.1 \times 10^3$ ) and L5 (mean  $1.8 \times 10^4$ , std.  $3.1 \times 10^3$ ) are overlapping significantly.

(B) Distribution of inhibitory neuron densities according to cortical layers for each subregion of the isocortex. For most of the cortical subregions, the density is constant from L2 to L5 and drops for L6.



Sup. Figure 6: **Resting state firing rate comparison for the different e-type populations of SSCtx-LL between the WBM1 and WBM2.**

Comparison of the mean instantaneous firing rate of the different e-type populations between the different versions of the whole mouse brain model in a resting state. The WBM1 (in dark orange) corresponds to the version presented in Erö [2], while the WBM2 (in blue) represents the refined version after the extension to the cell atlas (see 0) and the local connectivity refinement of Chapter 4.



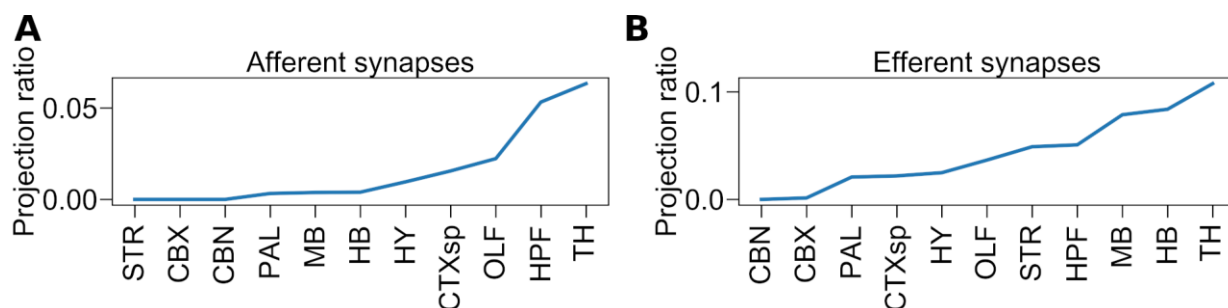
Sup. Figure 7: **Spiking activity of the excitatory and inhibitory populations of the whole mouse brain model.**

This figure shows the results of the WBM point neuron network simulated for 4 seconds while stimulated by a low background noise activity.

(A)(B) Comparison of the mean instantaneous firing rate of respectively the excitatory and inhibitory populations between the different versions of the whole mouse brain model. The WBM1 and WBM2 appear respectively in orange and blue.

(C) Spectrum analyses comparisons for the whole neuron population of the WBM1 and WBM2, shown in respectively orange and blue.

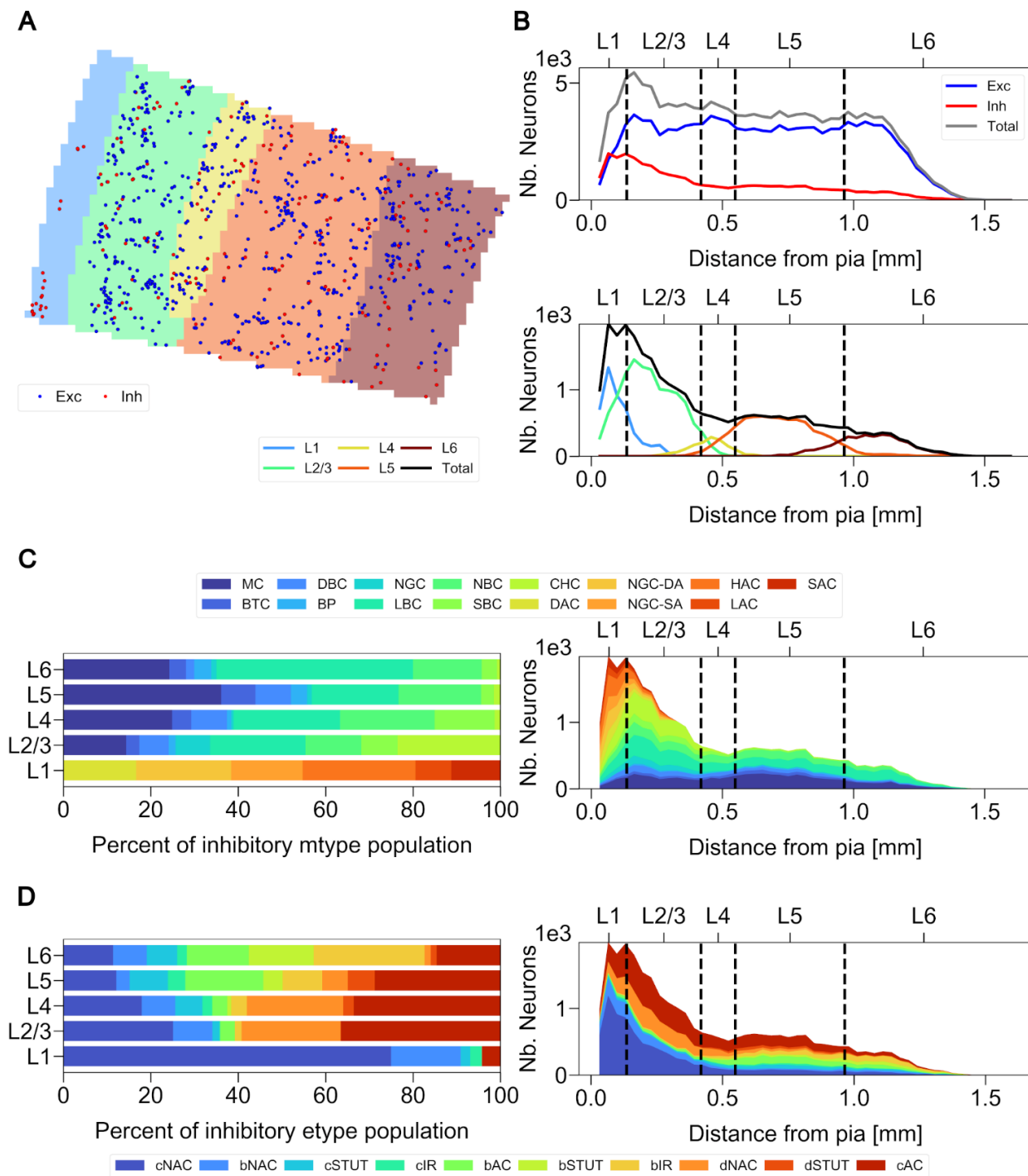
(D) Raster plot of the whole brain model spiking activity sorted by brain region. Each dot represents a spike, and its color corresponds to its region. The color coding matches the AIBS brain region color palette.



Sup. Figure 8: **Afferent and efferent connections of the SSCtx-LL.**

(A) Distribution of the connections coming from the rest of the brain towards the somatosensory cortex and (B) going out of the somatosensory cortex. The connections with the Isocortex are in majority and therefore not shown. Region shown: Thalamus (TH), Hippocampal formation (HPF), OLF (Olfactory areas), Cortical subplate (CTXsp), Hypothalamus (HY), Hindbrain (HB), Midbrain (MB), Pallidum (PAL), Cerebellar nuclei (CBN), Cerebellar cortex (CBX) and Striatum (STR).





Sup. Figure 9: **Neuron type distribution in the BBcAv1 Somatosensory cortex lower limb.**

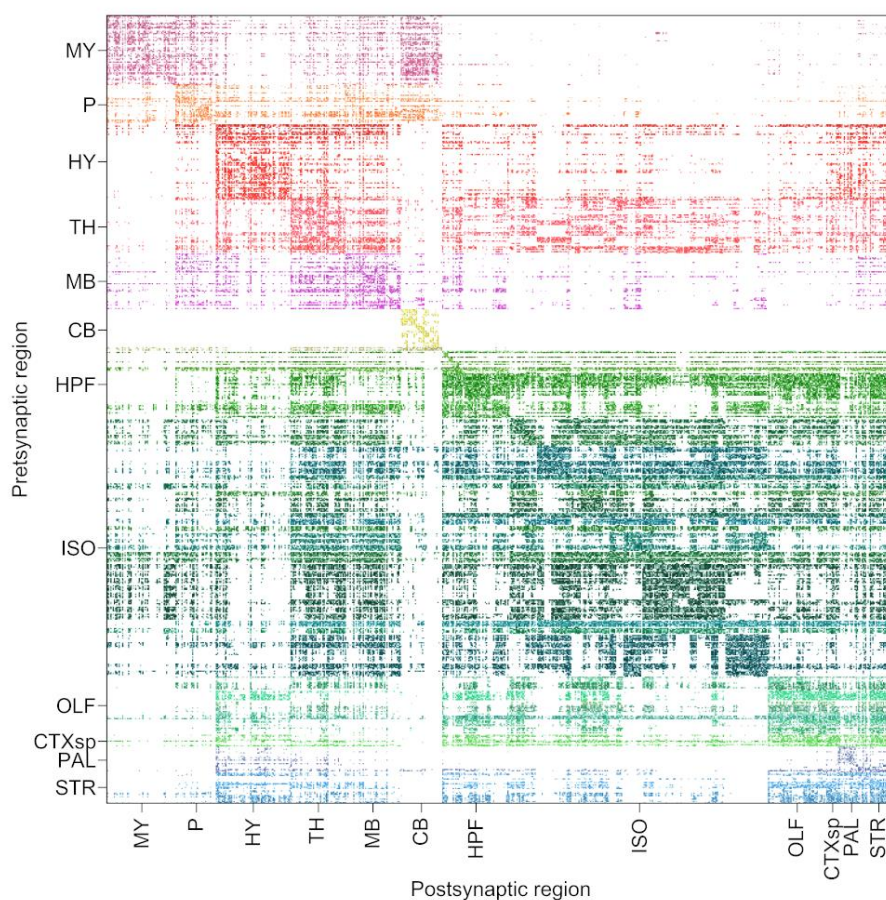
A. Coronal view of the BBcAv1 neuron distribution in the Somatosensory cortex lower limb region (SSCtx-LL). This panel shows the positions of distinct types of neurons in the different layers. Inhibitory and excitatory cells appear respectively in red and blue.

B. Distribution of neuron counts in SSCtx-LL according to their distance to the pia. The excitatory distribution is displayed with the blue line, the inhibitory population with the red line and the sum of the two is shown in gray. AV mean layer limits are represented as vertical dash lines.

C. Distribution of Inhibitory Morphological types (m-types) in SSCtx-LL according to their layer (left panel) and their distance from the pia (right panel). Each m-type population is represented with the same color in both panels. The left panel shows the proportion of the inhibitory m-types in each layer as percentage. The right panel shows the cumulative distribution of the inhibitory m-types according to their distance to the pia. AV mean layer limits are represented as vertical dash lines.

D. Distribution of Inhibitory Electrical types (e-types) in SSCtx-LL according to their layer (left panel) and their distance from the pia (right panel). The panels are analog to C.





Sup. Figure 10: **Generated neuron-to-neuron connectome of the refined whole mouse brain model.**

Cumulative combination of long-range and short-range connectivity. Each line and column of the matrix correspond to a brain region. Each pixel crossing a source and target region is either blank if there is no connection between the two regions or colored according to the source region. The colors correspond to the AIBS brain regions color palette. Acronyms correspond to the AV naming convention [34] (see Sup. Table 1). The transparency of the colored pixel is proportional to the total number of synapses between the pair of regions, in logarithmic scale.

## Supplementary tables

Abbrevia- tion	Full name		
Brain Regions			
ACA	Anterior Cingulate Area	OLF	Olfactory areas
ACAd	Anterior cingulate area, dorsal part	ORBI	Orbital area, lateral part
ACAv	Anterior cingulate area, ventral part	ORBm	Orbital area, medial part
Ald	Agranular insular area, dorsal part	ORBvl	Orbital area, ventrolateral part
Alp	Agranular insular area, posterior part	P	Pons
Alv	Agranular insular area, ventral part	PAL	Pallidum
AUDd	Dorsal auditory area	PERI	Perirhinal area
AUDp	primary auditory cortex	PL	Prelimbic area
AUDpo	Posterior auditory area	PTLp	Posterior parietal association areas
AUDv	Ventral auditory area	PV	parvalbumin
BCtx / SSp-bfd	Primary somatosensory area, barrel field	RSPagl	Retrosplenial area, lateral agranular part
CA1	Field CA1	RSPd	Retrosplenial area, dorsal part
CA3	Field CA3	RSPv	Retrosplenial area, ventral part
CB	Cerebellum	SSp	Primary somatosensory area
CBXpu	Cerebellar cortex, Purkinje layer	SSp-II	primary somatosensory cortex lower limb
CTXsp	Cortical subplate	SSp-m	Primary somatosensory area, mouth
DG	Dentate Gyrus	SSp-n	Primary somatosensory area, nose
ECT	Ectorhinal area	SSp-tr	Primary somatosensory area, trunk
FRP	Frontal pole, cerebral cortex	SSs	Supplemental somatosensory area
GU	Gustatory areas	STR	Striatum
HPF	Hippocampal Formation	SUB	Subiculum
HY	Hypothalamus	TEa	Temporal association areas
ILA	Infralimbic area	TH	Thalamus
ISO	Isocortex	VISal	Anterolateral visual area
MB	Midbrain	VISam	Anteromedial visual area
MOp	Primary motor area	VISC	Visceral area
MOs	Secondary motor area	VISl	Lateral visual area
MY	Medulla	VISp	Primary visual area
		VISpl	Posterolateral visual area
		VISpm	Posteromedial visual area

General		LAC	Large axon cells
AAV	Adeno-associated virus	LBC	Large basket cells
AIBS	Allen Institute for Brain Science	MC	Martinotti cells
AV	Annotation Volume	NBC	Nest basket cells
BBCA	Blue Brain Mouse Cell Atlas pipeline	NGC	Neurogliaform cells
BBP	Blue Brain Project	NGC-DA	Neurogliaform cell with dense axonal arborization
CCF	Common Coordinate Framework	NGC-SA	Neurogliaform cell with slender axonal arborization
EXC	Excitatory neurons	PC	Pyramidal cells
GABA	Gamma-Aminobutyric acid	SAC	Small axon cells
GAD	Glutamic acid decarboxylase	SBC	Small basket cells
GFP	Green Fluorescent Protein	SP	Star pyramidal cells
INH	Inhibitory neuron	SSC	Spiny stellate cells
IoU	Intersection over Union	TPC	Tufted pyramidal cells
ISH	<i>In situ</i> hybridization	UPC	Untufted pyramidal cells
LAMP5	Lysosomal associated membrane protein family member 5	Electrical types	
SST	Somatostatin	BAC	Burst accommodating
VIP	Vasoactive intestinal peptide	BIR	Burst Irregular
Morphological types		BNAC	Burst non-accommodating
BP	Bipolar cell	BSTUT	Burst stuttering
BPC	Bitufted pyramidal cells	CAC	Continuous accommodating
BTC	Bitufted cells	CADpyr	Continuous accommodating for pyramidal cells
CHC	Chandelier cells	CIR	Continuous irregular
DAC	Descending axon cells	CNAC	Continuous non-accommodating
DBC	Double bouquet cells	CSTUT	Continuous stuttering
HPC	Horizontal pyramidal cells	DNAC	Delayed non-accommodating
IPC	Inverted pyramidal cells	DSTUT	Delayed stuttering

Sup. Table 1: **Non standard abbreviations.**

		PV	SST	VIP	GAD67 exp. 1	GAD67 exp. 2	GAD67 combined
<b>Cerebellum</b>	<b>alpha</b>	455673	141748	79034	578316	578316	578316
	<b>R<sup>2</sup></b>	0,431	0,330	0,264	0,276	0,276	0,276
	<b>std</b>	28429,062	23931,201	16547,927	67953,049	67953,049	67953,049
<b>Isocortex</b>	<b>alpha</b>	296406	252262	231141	345848	345848	345848
	<b>R<sup>2</sup></b>	0,464	0,506	0,374	0,370	0,370	0,370
	<b>std</b>	9666,257	7804,143	10478,781	33750,389	33750,389	33750,389
<b>Rest of the brain</b>	<b>alpha</b>	118247	315793	115370	526450	526450	526450
	<b>R<sup>2</sup></b>	0,479	0,343	0,565	0,135	0,135	0,135
	<b>std</b>	4867,745	16968,697	5784,828	99696,489	99696,489	99696,489

Sup. Table 2: Results of the transfer function fitting.

# References

- [1] Blue Brain Project, “Blue Brain Portal,” *Blue Brain Portal*, 2022. <https://portal.bluebrain.epfl.ch/>
- [2] C. Erö, “Data-driven reconstruction of a point neuron mouse brain,” Lausanne, EPFL, 2018. Accessed: Mar. 15, 2022. [Online]. Available: <http://infoscience.epfl.ch/record/260465>
- [3] K. M. Stiefel and D. S. Brooks, “Why is There No Successful Whole Brain Simulation (Yet)?,” *Biol. Theory*, vol. 14, no. 2, pp. 122–130, Jun. 2019, doi: 10.1007/s13752-019-00319-5.
- [4] H. Markram *et al.*, “Reconstruction and Simulation of Neocortical Microcircuitry,” *Cell*, vol. 163, no. 2, pp. 456–492, Oct. 2015, doi: 10.1016/j.cell.2015.09.029.
- [5] A. Ecker *et al.*, “Data-driven integration of hippocampal CA1 synaptic physiology *in silico*,” *Hippocampus*, vol. 30, no. 11, pp. 1129–1145, Nov. 2020, doi: 10.1002/hipo.23220.
- [6] C. Eliasmith *et al.*, “A Large-Scale Model of the Functioning Brain,” *Science*, vol. 338, no. 6111, pp. 1202–1205, Nov. 2012, doi: 10.1126/science.1225266.
- [7] H. W. Dong, *Allen reference atlas: a digital color brain atlas of the C57Black/6J male mouse*. Hoboken, N.J.: Wiley, 2007.
- [8] C. Erö, M.-O. Gewaltig, D. Keller, and H. Markram, “A Cell Atlas for the Mouse Brain,” *Front. Neuroinformatics*, vol. 12, 2018, doi: 10.3389/fninf.2018.00084.
- [9] L. A. Jorgenson *et al.*, “The BRAIN Initiative: developing technology to catalyse neuroscience discovery,” *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 370, no. 1668, p. 20140164, May 2015, doi: 10.1098/rstb.2014.0164.
- [10] BRAIN Initiative Cell Census Network (BICCN) *et al.*, “A multimodal cell census and atlas of the mammalian primary motor cortex,” *Nature*, vol. 598, no. 7879, pp. 86–102, Oct. 2021, doi: 10.1038/s41586-021-03950-0.
- [11] R. Muñoz-Castañeda *et al.*, “Cellular anatomy of the mouse primary motor cortex,” *Nature*, vol. 598, no. 7879, pp. 159–166, Oct. 2021, doi: 10.1038/s41586-021-03970-w.
- [12] National Institutes of Health, “The BRAIN Initiative ®,” 2021. <https://braininitiative.nih.gov/>
- [13] S. Ramaswamy and H. Markram, “Anatomy and physiology of the thick-tufted layer 5 pyramidal neuron,” *Front. Cell. Neurosci.*, vol. 9, p. 233, Jun. 2015, doi: 10.3389/fncel.2015.00233.
- [14] S. Casali, E. Marenzi, C. Medini, C. Casellato, and E. D’Angelo, “Reconstruction and Simulation of a Scaffold Model of the Cerebellar Network,” *Front. Neuroinformatics*, vol. 13, 2019, doi: 10.3389/fninf.2019.00037.
- [15] M. S. Todtenkopf, S. L. Vincent, and F. M. Benes, “A cross-study meta-analysis and three-dimensional comparison of cell counting in the anterior cingulate cortex of schizophrenic and bipolar brain,” *Schizophr. Res.*, vol. 73, no. 1, pp. 79–89, Feb. 2005, doi: 10.1016/j.schres.2004.08.018.
- [16] Y. Kim *et al.*, “Brain-wide Maps Reveal Stereotyped Cell-Type-Based Cortical Architecture and Subcortical Sexual Dimorphism,” *Cell*, vol. 171, no. 2, pp. 456–469.e22, Oct. 2017, doi: 10.1016/j.cell.2017.09.020.
- [17] D. Keller, C. Erö, and H. Markram, “Cell Densities in the Mouse Brain: A Systematic Review,” *Front. Neuroanat.*, vol. 12, p. 83, Oct. 2018, doi: 10.3389/fnana.2018.00083.
- [18] S. Lang, V. J. Dercksen, B. Sakmann, and M. Oberlaender, “Simulation of signal flow in 3D reconstructions of an anatomically realistic neural network in rat vibrissa cortex,” *Neural Netw.*, vol. 24, no. 9, pp. 998–1011, Nov. 2011, doi: 10.1016/j.neunet.2011.06.013.
- [19] M. Lindahl and J. Hellgren Kotaleski, “Untangling Basal Ganglia Network Dynamics and Function: Role of Dopamine Depletion and Inhibition Investigated in a Spiking Network Model,” *eneuro*, vol. 3, no. 6, p. ENEURO.0156-16.2016, Nov. 2016, doi: 10.1523/ENEURO.0156-16.2016.
- [20] Y. N. Billeh *et al.*, “Systematic Integration of Structural and Functional Data into Multi-scale Models of Mouse Primary Visual Cortex,” *Neuron*, vol. 106, no. 3, Art. no. 3, May 2020, doi: 10.1016/j.neuron.2020.01.040.
- [21] E. S. Lein *et al.*, “Genome-wide atlas of gene expression in the adult mouse brain,” *Nature*, vol. 445, no. 7124, pp. 168–176, Jan. 2007, doi: 10.1038/nature05453.
- [22] S. Herculano-Houzel *et al.*, “Updated Neuronal Scaling Rules for the Brains of Glires (Rodents/Lagomorphs),” *Brain. Behav. Evol.*, vol. 78, no. 4, pp. 302–314, 2011, doi: 10.1159/000330825.
- [23] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, “Tracking Whole-Brain Connectivity Dynamics in the Resting State,” *Cereb. Cortex*, vol. 24, no. 3, pp. 663–676, Mar. 2014, doi: 10.1093/cercor/bhs352.

- [24] B. R. White, A. Q. Bauer, A. Z. Snyder, B. L. Schlaggar, J.-M. Lee, and J. P. Culver, "Imaging of Functional Connectivity in the Mouse Brain," *PLoS ONE*, vol. 6, no. 1, p. e16322, Jan. 2011, doi: 10.1371/journal.pone.0016322.
- [25] M. P. Vanni, A. W. Chan, M. Balbi, G. Silasi, and T. H. Murphy, "Mesoscale Mapping of Mouse Cortex Reveals Frequency-Dependent Cycling between Distinct Macroscale Functional Modules," *J. Neurosci.*, vol. 37, no. 31, pp. 7513–7533, Aug. 2017, doi: 10.1523/JNEUROSCI.3560-16.2017.
- [26] M. H. Mohajerani *et al.*, "Spontaneous cortical activity alternates between motifs defined by regional axonal projections," *Nat. Neurosci.*, vol. 16, no. 10, pp. 1426–1435, Oct. 2013, doi: 10.1038/nn.3499.
- [27] M. W. Reimann, J. G. King, E. B. Muller, S. Ramaswamy, and H. Markram, "An algorithm to predict the connectome of neural microcircuits," *Front. Comput. Neurosci.*, vol. 9, Oct. 2015, doi: 10.3389/fncom.2015.00120.
- [28] J. Winnubst *et al.*, "Reconstruction of 1,000 Projection Neurons Reveals New Cell Types and Organization of Long-Range Connectivity in the Mouse Brain," *Cell*, vol. 179, no. 1, pp. 268–281.e13, Sep. 2019, doi: 10.1016/j.cell.2019.07.042.
- [29] S. W. Oh *et al.*, "A mesoscale connectome of the mouse brain," *Nature*, vol. 508, no. 7495, pp. 207–214, Apr. 2014, doi: 10.1038/nature13186.
- [30] J. A. Harris *et al.*, "The organization of intracortical connections by layer and cell class in the mouse brain," *Neuroscience*, preprint, Apr. 2018. doi: 10.1101/292961.
- [31] B. Zingg *et al.*, "Neural Networks of the Mouse Neocortex," *Cell*, vol. 156, no. 5, pp. 1096–1111, Feb. 2014, doi: 10.1016/j.cell.2014.02.023.
- [32] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii, "Structural Properties of the *Caenorhabditis elegans* Neuronal Network," *PLoS Comput. Biol.*, vol. 7, no. 2, p. e1001066, Feb. 2011, doi: 10.1371/journal.pcbi.1001066.
- [33] L. Kuan *et al.*, "Neuroinformatics of the Allen Mouse Brain Connectivity Atlas," *Methods*, vol. 73, pp. 4–17, Feb. 2015, doi: 10.1016/j.ymeth.2014.12.013.
- [34] Q. Wang *et al.*, "The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas," *Cell*, vol. 181, no. 4, pp. 936–953.e20, May 2020, doi: 10.1016/j.cell.2020.04.007.
- [35] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *NeuroImage*, vol. 54, no. 3, pp. 2033–2044, Feb. 2011, doi: 10.1016/j.neuroimage.2010.09.025.
- [36] J. Krepl *et al.*, "Supervised Learning with Perceptual Similarity for Multimodal Gene Expression Registration of a Mouse Brain Atlas," 2021.
- [37] Allen Institute for Brain Science, "CCFv2 Adult Mouse Annotation Atlas," 2015. [http://download.alleninstitute.org/informatics-archive/current-release/mouse\\_ccf/annotation/ccf\\_2015/](http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/annotation/ccf_2015/)
- [38] Allen Institute for Brain Science, "Adult Mouse Allen Reference Atlas," 2015. [http://download.alleninstitute.org/informatics-archive/current-release/mouse\\_ccf/ara\\_nissl/](http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/ara_nissl/)
- [39] Allen Institute for Brain Science, "CCFv3 Adult Mouse Annotation Atlas," 2017. [http://download.alleninstitute.org/informatics-archive/current-release/mouse\\_ccf/annotation/ccf\\_2017/](http://download.alleninstitute.org/informatics-archive/current-release/mouse_ccf/annotation/ccf_2017/)
- [40] I. Gourfinkel-An *et al.*, "Changes in GAD67 mRNA expression evidenced by in situ hybridization in the brain of R6/2 transgenic mice," *J. Neurochem.*, vol. 86, no. 6, pp. 1369–1378, 2003, doi: <https://doi.org/10.1046/j.1471-4159.2003.01916.x>.
- [41] S. Lefort, C. Tómm, J.-C. Floyd Sarria, and C. C. H. Petersen, "The Excitatory Neuronal Network of the C2 Barrel Column in Mouse Primary Somatosensory Cortex," *Neuron*, vol. 61, no. 2, pp. 301–316, Jan. 2009, doi: 10.1016/j.neuron.2008.12.020.
- [42] X. Xu, K. D. Roby, and E. M. Callaway, "Immunohistochemical characterization of inhibitory mouse cortical neurons: Three chemically distinct classes of inhibitory cells," *J. Comp. Neurol.*, vol. 518, no. 3, pp. 389–404, Feb. 2010, doi: 10.1002/cne.22229.
- [43] S. Jinno, Y. Aika, T. Fukuda, and T. Kosaka, "Quantitative analysis of GABAergic neurons in the mouse hippocampus, with optical disector using confocal laser scanning microscope," *Brain Res.*, vol. 814, no. 1–2, pp. 55–70, Dec. 1998, doi: 10.1016/S0006-8993(98)01075-0.
- [44] S. Herculano-Houzel, C. Watson, and G. Paxinos, "Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones," *Front. Neuroanat.*, vol. 7, p. 35, 2013, doi: 10.3389/fnana.2013.00035.
- [45] D. M. Young *et al.*, "Constructing and optimizing 3D atlases from 2D data with application to the developing mouse brain," *eLife*, vol. 10, p. e61408, Feb. 2021, doi: 10.7554/eLife.61408.
- [46] G. Paxinos and K. B. J. Franklin, *Paxinos and Franklin's The mouse brain in stereotaxic coordinates*, Fifth edition. London San Diego Cambridge; MA Kidlington, Oxford: Elsevier, Academic Press, 2019.
- [47] U. Chon, D. J. Vanselow, K. C. Cheng, and Y. Kim, "Enhanced and unified anatomical labeling for a common mouse brain atlas," *Nat. Commun.*, vol. 10, no. 1, p. 5067, Dec. 2019, doi: 10.1038/s41467-019-13057-w.
- [48] C. O'Reilly, E. Iavarone, J. Yi, and S. L. Hill, "Rodent Somatosensory Thalamocortical Circuitry: Neurons, Synapses, and Connectivity," *LIFE SCIENCES*, preprint, Jul. 2020. doi: 10.20944/preprints202007.0053.v1.
- [49] H. Hu, J. Gan, and P. Jonas, "Fast-spiking, parvalbumin + GABAergic interneurons: From cellular design to microcircuit function," *Science*, vol. 345, no. 6196, p. 1255263, Aug. 2014, doi: 10.1126/science.1255263.
- [50] N. R. Wall, M. De La Parra, J. M. Sorokin, H. Taniguchi, Z. J. Huang, and E. M. Callaway, "Brain-Wide Maps of Synaptic Input to Cortical Interneurons," *J. Neurosci.*, vol. 36, no. 14, pp. 4000–4009, Apr. 2016, doi: 10.1523/JNEUROSCI.3967-15.2016.

- [51] B. Schmalbach *et al.*, "Age-dependent loss of parvalbumin-expressing hippocampal interneurons in mice deficient in CHL1, a mental retardation and schizophrenia susceptibility gene," *J. Neurochem.*, vol. 135, no. 4, pp. 830–844, 2015, doi: <https://doi.org/10.1111/jnc.13284>.
- [52] L. Fasulo *et al.*, "ProNGF Drives Localized and Cell Selective Parvalbumin Interneuron and Perineuronal Net Depletion in the Dentate Gyrus of Transgenic Mice," *Front. Mol. Neurosci.*, vol. 10, Feb. 2017, doi: 10.3389/fnmol.2017.00020.
- [53] E. Lauber, F. Filice, and B. Schwaller, "Dysregulation of Parvalbumin Expression in the Cntnap2<sup>-/-</sup> Mouse Model of Autism Spectrum Disorder," *Front. Mol. Neurosci.*, vol. 11, p. 262, Aug. 2018, doi: 10.3389/fnmol.2018.00262.
- [54] A. Pirone *et al.*, "Social Stimulus Causes Aberrant Activation of the Medial Prefrontal Cortex in a Mouse Model With Autism-Like Behaviors," *Front. Synaptic Neurosci.*, vol. 10, p. 35, Oct. 2018, doi: 10.3389/fnsyn.2018.00035.
- [55] B. Ramos *et al.*, "Early neuropathology of somatostatin/NPY GABAergic cells in the hippocampus of a PS1×APP transgenic model of Alzheimer's disease," *Neurobiol. Aging*, vol. 27, no. 11, pp. 1658–1672, Nov. 2006, doi: 10.1016/j.neurobiolaging.2005.09.022.
- [56] L. Trujillo-Estrada *et al.*, "Early Neuronal Loss and Axonal/Presynaptic Damage is Associated with Accelerated Amyloid- $\beta$  Accumulation in A $\beta$ PP/PS1 Alzheimer's Disease Mice Subiculum," *J. Alzheimers Dis.*, vol. 42, no. 2, pp. 521–541, Aug. 2014, doi: 10.3233/JAD-140495.
- [57] E. Sanchez-Mejias *et al.*, "Distinct disease-sensitive GABAergic neurons in the perirhinal cortex of Alzheimer's mice and patients," *Brain Pathol.*, vol. 30, no. 2, pp. 345–363, Mar. 2020, doi: 10.1111/bpa.12785.
- [58] A. Prönneke, B. Scheuer, R. J. Wagener, M. Möck, M. Witte, and J. F. Staiger, "Characterizing VIP Neurons in the Barrel Cortex of VIPcre/tomato Mice Reveals Layer-Specific Differences," *Cereb. Cortex*, vol. 25, no. 12, pp. 4854–4868, Dec. 2015, doi: 10.1093/cercor/bhv202.
- [59] S. Passemard, C. A. Rabat, P. Gressens, and V. Lelièvre, "Chapter 128 - VIP," in *Handbook of Biologically Active Peptides*, 2nd ed., Cambridge, Massachusetts: Elsevier, 2013, pp. 966–974. doi: 10.1016/B978-0-12-385095-9.00128-7.
- [60] Z. J. Huang and A. Paul, "The diversity of GABAergic neurons and neural communication elements," *Nat. Rev. Neurosci.*, vol. 20, no. 9, pp. 563–572, Sep. 2019, doi: 10.1038/s41583-019-0195-4.
- [61] G. A. Ascoli, D. E. Donohue, and M. Halavi, "NeuroMorpho.Org: A Central Resource for Neuronal Morphologies," *J. Neurosci.*, vol. 27, no. 35, pp. 9247–9251, Aug. 2007, doi: 10.1523/JNEUROSCI.2055-07.2007.
- [62] S. J. Tripathy, J. Savitskaya, S. D. Burton, N. N. Urban, and R. C. Gerkin, "NeuroElectro: a window to the world's neuron electrophysiology data," *Front. Neuroinformatics*, vol. 8, Apr. 2014, doi: 10.3389/fninf.2014.00040.
- [63] Y. Roussel *et al.*, "Mapping of morpho-electric features to molecular identity of cortical inhibitory neurons," *Neuroscience*, preprint, Nov. 2021. doi: 10.1101/2021.11.24.469815.
- [64] Allen Institute for Brain Science, "ALLEN Mouse Brain Atlas. Technical white paper: Informatics data processing," *Allen Institute for Brain Science*, 2011. <http://help.brain-map.org/download/attachments/2818169/InformaticsDataProcessing.pdf>
- [65] Allen Institute for Brain Science, "ISH Data Allen Mouse Brain Atlas," 2011. <http://mouse.brain-map.org/search/index>
- [66] A. Zeisel *et al.*, "Molecular Architecture of the Mouse Nervous System," *Cell*, vol. 174, no. 4, pp. 999–1014.e22, Aug. 2018, doi: 10.1016/j.cell.2018.06.021.
- [67] H. Asada *et al.*, "Cleft palate and decreased brain -aminobutyric acid in mice lacking the 67-kDa isoform of glutamic acid decarboxylase," *Proc. Natl. Acad. Sci.*, vol. 94, no. 12, pp. 6496–6499, Jun. 1997, doi: 10.1073/pnas.94.12.6496.
- [68] S. F. Kash *et al.*, "Epilepsy in mice deficient in the 65-kDa isoform of glutamic acid decarboxylase," *Proc. Natl. Acad. Sci.*, vol. 94, no. 25, pp. 14060–14065, Dec. 1997, doi: 10.1073/pnas.94.25.14060.
- [69] M. R. Celio and C. W. Heizmann, "Calcium-binding protein parvalbumin as a neuronal marker," *Nature*, vol. 293, no. 5830, pp. 300–302, Sep. 1981, doi: 10.1038/293300a0.
- [70] P. C. Emson, "Somatostatin and Receptors," in *Encyclopedia of Neuroscience*, L. R. Squire, Ed. Oxford: Academic Press, 2009, pp. 121–127. doi: 10.1016/B978-008045046-9.01473-X.
- [71] M. Celio, "Parvalbumin in most gamma-aminobutyric acid-containing neurons of the rat cerebral cortex," *Science*, vol. 231, no. 4741, pp. 995–997, Feb. 1986, doi: 10.1126/science.3945815.
- [72] B. Tasic *et al.*, "Shared and distinct transcriptomic cell types across neocortical areas," *Nature*, vol. 563, no. 7729, pp. 72–78, Nov. 2018, doi: 10.1038/s41586-018-0654-5.
- [73] A. Irintchev, A. Rollenhagen, E. Troncoso, J. Z. Kiss, and M. Schachner, "Structural and Functional Aberrations in the Cerebral Cortex of Tenascin-C Deficient Mice," *Cereb. Cortex*, vol. 15, no. 7, Art. no. 7, Jul. 2005, doi: 10.1093/cercor/bhh195.
- [74] M. I. Ransome and A. M. Turnley, "Analysis of neuronal subpopulations in mice over-expressing suppressor of cytokine signaling-2," *Neuroscience*, vol. 132, no. 3, pp. 673–687, Jan. 2005, doi: 10.1016/j.neuroscience.2004.12.041.
- [75] J. Neddens and A. Buonanno, "Selective populations of hippocampal interneurons express ErbB4 and their number and distribution is altered in ErbB4 knockout mice," *Hippocampus*, vol. 20, no. 6, pp. 724–744, 2010, doi: <https://doi.org/10.1002/hipo.20675>.
- [76] J. S. Schmid *et al.*, "Heterozygosity for the mutated X-chromosome-linked L1 cell adhesion molecule gene leads to increased numbers of neurons and enhanced metabolism in the forebrain of female carrier mice," *Brain Struct. Funct.*, vol. 218, no. 6, pp. 1375–1390, Nov. 2013, doi: 10.1007/s00429-012-0463-9.

- [77] L.-C. Han *et al.*, "The Effect of Sevoflurane Inhalation on Gabaergic Neurons Activation: Observation on the GAD67-GFP Knock-In Mouse," *Anat. Rec. Adv. Integr. Anat. Evol. Biol.*, vol. 293, no. 12, pp. 2114–2122, Dec. 2010, doi: 10.1002/ar.21113.
- [78] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Syst. Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979, doi: 10.1109/TSMC.1979.4310076.
- [79] P. Virtanen *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nat. Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.
- [80] C. R. Gerfen and C. J. Wilson, "Chapter II The basal ganglia," in *Handbook of Chemical Neuroanatomy*, vol. 12, Elsevier, 1996, pp. 371–468. doi: 10.1016/S0924-8196(96)80004-2.
- [81] N. W. Gouwens *et al.*, "Classification of electrophysiological and morphological neuron types in the mouse visual cortex," *Nat. Neurosci.*, vol. 22, no. 7, pp. 1182–1195, Jul. 2019, doi: 10.1038/s41593-019-0417-0.
- [82] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," 2011, doi: 10.48550/ARXIV.1109.2378.
- [83] W. Lange, "Cell number and cell density in the cerebellar cortex of man and some other mammals," *Cell Tissue Res.*, vol. 157, no. 1, Art. no. 1, Mar. 1975, doi: 10.1007/BF00223234.
- [84] J. A. Förster, "Quantitative morphological analysis of the neostriatum and the cerebellum of tenascin-C deficient mice," *Quantitative morphologische Analysen des Neostriatums und des Cerebellums der Tenascin-C defizienten Maus*, 2008, Accessed: Jan. 18, 2021. [Online]. Available: <https://ediss.sub.uni-hamburg.de/handle/ediss/2354>
- [85] G. Sekerková, M. Watanabe, M. Martina, and E. Mugnaini, "Differential distribution of phospholipase C beta isoforms and diacylglycerol kinase-beta in rodents cerebella corroborates the division of unipolar brush cells into two major subtypes," *Brain Struct. Funct.*, vol. 219, no. 2, pp. 719–749, Mar. 2014, doi: 10.1007/s00429-013-0531-9.
- [86] S. Dieudonné and A. Dumoulin, "Serotonin-Driven Long-Range Inhibitory Connections in the Cerebellar Cortex," *J. Neurosci.*, vol. 20, no. 5, pp. 1837–1848, Mar. 2000, doi: 10.1523/JNEUROSCI.20-05-01837.2000.
- [87] H. S. Meyer *et al.*, "Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5A," *Proc. Natl. Acad. Sci.*, vol. 108, no. 40, pp. 16807–16812, Oct. 2011, doi: 10.1073/pnas.1113648108.
- [88] J. Waider, F. Proft, G. Langlhofer, E. Asan, K.-P. Lesch, and L. Gutknecht, "GABA concentration and GABAergic neuron populations in limbic areas are differentially altered by brain serotonin deficiency in Tph2 knockout mice," *Histochem. Cell Biol.*, vol. 139, no. 2, pp. 267–281, Feb. 2013, doi: 10.1007/s00418-012-1029-x.
- [89] S. Trifonov, Y. Yamashita, M. Kase, M. Maruyama, and T. Sugimoto, "Glutamic acid decarboxylase 1 alternative splicing isoforms: characterization, expression and quantification in the mouse brain," *BMC Neurosci.*, vol. 15, no. 1, p. 114, Oct. 2014, doi: 10.1186/1471-2202-15-114.
- [90] X. Wang *et al.*, "Three-dimensional intact-tissue sequencing of single-cell transcriptional states," *Science*, vol. 361, no. 6400, p. eaat5691, Jul. 2018, doi: 10.1126/science.aat5691.
- [91] V. Szabolcsi and M. R. Celio, "De novo expression of parvalbumin in ependymal cells in response to brain injury promotes ependymal remodeling and wound repair," *Glia*, vol. 63, no. 4, pp. 567–594, 2015, doi: <https://doi.org/10.1002/glia.22768>.
- [92] S. Lee, J. Hjerling-Leffler, E. Zagha, G. Fishell, and B. Rudy, "The Largest Group of Superficial Neocortical GABAergic Interneurons Expresses Ionotropic Serotonin Receptors," *J. Neurosci.*, vol. 30, no. 50, pp. 16796–16808, Dec. 2010, doi: 10.1523/JNEUROSCI.1869-10.2010.
- [93] S. Parrish-Aungst, M. T. Shipley, F. Erdelyi, G. Szabo, and A. C. Puche, "Quantitative analysis of neuronal diversity in the mouse olfactory bulb," *J. Comp. Neurol.*, vol. 501, no. 6, pp. 825–836, Apr. 2007, doi: 10.1002/cne.21205.
- [94] N. Suzuki and J. M. Bekkers, "Inhibitory neurons in the anterior piriform cortex of the mouse: Classification using molecular markers," *J. Comp. Neurol.*, vol. 518, no. 10, pp. 1670–1687, May 2010, doi: 10.1002/cne.22295.
- [95] J. Tepper, F. Tecuapetla, T. Koos, and O. Ibanez-Sandoval, "Heterogeneity and Diversity of Striatal GABAergic Interneurons," *Front. Neuroanat.*, vol. 4, 2010, Accessed: Jun. 13, 2022. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fnana.2010.00150>
- [96] L. Kanari *et al.*, "Objective Classification of Neocortical Pyramidal Cells." *bioRxiv*, p. 349977, Jun. 19, 2018. doi: 10.1101/349977.
- [97] T. C. Potjans and M. Diesmann, "The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model," *Cereb. Cortex*, vol. 24, no. 3, pp. 785–806, Mar. 2014, doi: 10.1093/cercor/bhs358.
- [98] H. von Gersdorff, R. Schneggenburger, S. Weis, and E. Neher, "Presynaptic Depression at a Calyx Synapse: The Small Contribution of Metabotropic Glutamate Receptors," *J. Neurosci.*, vol. 17, no. 21, pp. 8137–8146, Nov. 1997, doi: 10.1523/JNEUROSCI.17-21-08137.1997.
- [99] M. W. Reimann, A.-L. Horlemann, S. Ramaswamy, E. B. Muller, and H. Markram, "Morphological Diversity Strongly Constrains Synaptic Connectivity and Plasticity," *Cereb. Cortex*, vol. 27, no. 9, pp. 4570–4585, Sep. 2017, doi: 10.1093/cercor/bhx150.
- [100] C. Rössert *et al.*, "Automated point-neuron simplification of data-driven microcircuit models," 2016, doi: 10.48550/ARXIV.1604.00087.
- [101] M.-O. Gewaltig and M. Diesmann, "NEST (NEural Simulation Tool)," *Scholarpedia*, vol. 2, no. 4, p. 1430, 2007, doi: 10.4249/scholarpedia.1430.



- [102] N. Urbain, P. A. Salin, P.-A. Libourel, J.-C. Comte, L. J. Gentet, and C. C. H. Petersen, "Whisking-Related Changes in Neuronal Firing and Membrane Potential Dynamics in the Somatosensory Thalamus of Awake Mice," *Cell Rep.*, vol. 13, no. 4, pp. 647–656, Oct. 2015, doi: 10.1016/j.celrep.2015.09.029.
- [103] R. Brette and W. Gerstner, "Adaptive Exponential Integrate-and-Fire Model as an Effective Description of Neuronal Activity," *J. Neurophysiol.*, vol. 94, no. 5, pp. 3637–3642, Nov. 2005, doi: 10.1152/jn.00686.2005.
- [104] R. Naud, N. Marcille, C. Clopath, and W. Gerstner, "Firing patterns in the adaptive exponential integrate-and-fire model," *Biol. Cybern.*, vol. 99, no. 4–5, pp. 335–347, Nov. 2008, doi: 10.1007/s00422-008-0264-7.
- [105] M. V. Tsodyks and H. Markram, "The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability," *Proc. Natl. Acad. Sci.*, vol. 94, no. 2, pp. 719–723, Jan. 1997, doi: 10.1073/pnas.94.2.719.
- [106] P. A. Simmons and A. L. Pearlman, "Receptive-field properties of transcallosal visual cortical neurons in the normal and reeler mouse," *J. Neurophysiol.*, vol. 50, no. 4, pp. 838–848, Oct. 1983, doi: 10.1152/jn.1983.50.4.838.
- [107] M. Nawrot, A. Aertsen, and S. Rotter, "Single-trial estimation of neuronal firing rates: From single-neuron spike trains to population activity," *J. Neurosci. Methods*, vol. 94, no. 1, pp. 81–92, Dec. 1999, doi: 10.1016/S0165-0270(99)00127-2.
- [108] G. R. Holt, W. R. Softky, C. Koch, and R. J. Douglas, "Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons," *J. Neurophysiol.*, vol. 75, no. 5, pp. 1806–1814, May 1996, doi: 10.1152/jn.1996.75.5.1806.
- [109] U. Fano, "Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions," *Phys. Rev.*, vol. 72, no. 1, pp. 26–29, Jul. 1947, doi: 10.1103/PhysRev.72.26.
- [110] X.-J. Wang, "Calcium Coding and Adaptive Temporal Computation in Cortical Pyramidal Neurons," *J. Neurophysiol.*, vol. 79, no. 3, pp. 1549–1566, Mar. 1998, doi: 10.1152/jn.1998.79.3.1549.
- [111] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*. Springer Verlag, 2014.
- [112] M. P. Nawrot, C. Boucsein, V. Rodriguez Molina, A. Riehle, A. Aertsen, and S. Rotter, "Measurement of variability dynamics in cortical spike trains," *J. Neurosci. Methods*, vol. 169, no. 2, pp. 374–390, Apr. 2008, doi: 10.1016/j.jneumeth.2007.10.013.
- [113] G. T. Neske, "The Slow Oscillation in Cortical and Thalamic Networks: Mechanisms and Functions," *Front. Neural Circuits*, vol. 9, Jan. 2016, doi: 10.3389/fncir.2015.00088.
- [114] H. S. Meyer *et al.*, "Cell Type-Specific Thalamic Innervation in a Column of Rat Vibrissa Cortex," *Cereb. Cortex*, vol. 20, no. 10, pp. 2287–2303, Oct. 2010, doi: 10.1093/cercor/bhq069.
- [115] M. L. Hines and N. T. Carnevale, "The NEURON Simulation Environment," *Neural Comput.*, vol. 9, no. 6, pp. 1179–1209, Aug. 1997, doi: 10.1162/neco.1997.9.6.1179.
- [116] M. W. Reimann, M. Gevaert, Y. Shi, H. Lu, H. Markram, and E. Muller, "A null model of the mouse whole-neocortex micro-connectome," *Nat. Commun.*, vol. 10, no. 1, p. 3903, Dec. 2019, doi: 10.1038/s41467-019-11630-x.
- [117] I. Ferezou, F. Haiss, L. J. Gentet, R. Aronoff, B. Weber, and C. C. H. Petersen, "Spatiotemporal Dynamics of Cortical Sensorimotor Integration in Behaving Mice," *Neuron*, vol. 56, no. 5, pp. 907–923, Dec. 2007, doi: 10.1016/j.neuron.2007.10.007.
- [118] C. Zhang *et al.*, "A platform for stereological quantitative analysis of the brain-wide distribution of type-specific neurons," *Sci. Rep.*, vol. 7, no. 1, Art. no. 1, Dec. 2017, doi: 10.1038/s41598-017-14699-w.
- [119] S. Jinno and T. Kosaka, "Cellular architecture of the mouse hippocampus: A quantitative aspect of chemically defined GABAergic neurons with stereology," *Neurosci. Res.*, vol. 56, no. 3, Art. no. 3, Nov. 2006, doi: 10.1016/j.neures.2006.07.007.
- [120] C. Zhao, B. Eisinger, and S. C. Gammie, "Characterization of GABAergic Neurons in the Mouse Lateral Septum: A Double Fluorescence In Situ Hybridization and Immunohistochemical Study Using Tyramide Signal Amplification," *PLoS ONE*, vol. 8, no. 8, Art. no. 8, Aug. 2013, doi: 10.1371/journal.pone.0073750.
- [121] G. Hafner *et al.*, "Circuits in the absence of cortical layers: increased callosal connectivity in reeler mice revealed by brain-wide input mapping of VIP neurons in barrel cortex," *bioRxiv*, p. 2020.04.19.048868, Apr. 2020, doi: 10.1101/2020.04.19.048868.
- [122] J. Gotts, L. Atkinson, I. J. Edwards, Y. Yanagawa, S. A. Deuchars, and J. Deuchars, "Co-expression of GAD67 and choline acetyltransferase reveals a novel neuronal phenotype in the mouse medulla oblongata," *Auton. Neurosci.*, vol. 193, pp. 22–30, Dec. 2015, doi: 10.1016/j.autneu.2015.05.003.
- [123] P. D. Whissell, J. D. Cajanding, N. Fogel, and J. C. Kim, "Comparative density of CCK- and PV-GABA cells within the cortex and hippocampus," *Front. Neuroanat.*, vol. 9, 2015, doi: 10.3389/fnana.2015.00124.
- [124] M. W. Pitts *et al.*, "Deletion of Selenoprotein M Leads to Obesity without Cognitive Deficits \*," *J. Biol. Chem.*, vol. 288, no. 36, Art. no. 36, Sep. 2013, doi: 10.1074/jbc.M113.471235.
- [125] H. Yamanaka, Y. Yanagawa, and K. Obata, "Development of stellate and basket cells and their apoptosis in mouse cerebellar cortex," *Neurosci. Res.*, vol. 50, no. 1, Art. no. 1, Sep. 2004, doi: 10.1016/j.neures.2004.06.008.
- [126] Z. Almási, C. Dávid, M. Witte, and J. F. Staiger, "Distribution Patterns of Three Molecularly Defined Classes of GABAergic Neurons Across Columnar Compartments in Mouse Barrel Cortex," *Front. Neuroanat.*, vol. 13, 2019, doi: 10.3389/fnana.2019.00045.
- [127] G. Calfa, W. Li, J. M. Rutherford, and L. Pozzo-Miller, "Excitation/Inhibition Imbalance and Impaired Synaptic Inhibition in Hippocampal Area CA3 of Mecp2 Knockout Mice," *Hippocampus*, vol. 25, no. 2, pp. 159–168, Feb. 2015, doi: 10.1002/hipo.22360.
- [128] I. Moreno-Gonzalez *et al.*, "Extracellular Amyloid- $\beta$  and Cytotoxic Glial Activation Induce Significant Entorhinal Neuron Loss in Young PS1M146L/APP751SL Mice," *J. Alzheimers Dis.*, vol. 18, no. 4, Art. no. 4, Nov. 2009, doi: 10.3233/JAD-2009-1192.

- [129] M. Ono, Y. Yanagawa, and K. Koyano, "GABAergic neurons in inferior colliculus of the GAD67-GFP knock-in mouse: Electrophysiological and morphological properties," *Neurosci. Res.*, vol. 51, no. 4, Art. no. 4, Apr. 2005, doi: 10.1016/j.neures.2004.12.019.
- [130] P. Arcelli, C. Frassoni, M. C. Regondi, S. D. Biasi, and R. Spreafico, "GABAergic Neurons in Mammalian Thalamus: A Marker of Thalamic Complexity?," *Brain Res. Bull.*, vol. 42, no. 1, Art. no. 1, Jan. 1997, doi: 10.1016/S0361-9230(96)00107-4.
- [131] N. Tamamaki, Y. Yanagawa, R. Tomioka, J.-I. Miyazaki, K. Obata, and T. Kaneko, "Green fluorescent protein expression and colocalization with calretinin, parvalbumin, and somatostatin in the GAD67-GFP knock-in mouse," *J. Comp. Neurol.*, vol. 467, no. 1, Art. no. 1, Dec. 2003, doi: 10.1002/cne.10905.
- [132] T. A. Seabrook, T. E. Krahe, G. Govindaiah, and W. Guido, "Interneurons in the mouse visual thalamus maintain a high degree of retinal convergence throughout postnatal development," *Neural Develop.*, vol. 8, no. 1, Art. no. 1, 2013, doi: 10.1186/1749-8104-8-24.
- [133] G. Hou, A. G. Smith, and Z.-W. Zhang, "Lack of Intrinsic GABAergic Connections in the Thalamic Reticular Nucleus of the Mouse," *J. Neurosci.*, vol. 36, no. 27, Art. no. 27, Jul. 2016, doi: 10.1523/JNEUROSCI.0607-16.2016.
- [134] Y. Gonchar, Q. Wang, and A. H. Burkhalter, "Multiple distinct subtypes of GABAergic neurons in mouse visual cortex identified by triple immunostaining," *Front. Neuroanat.*, vol. 2, 2008, doi: 10.3389/neuro.05.003.2007.
- [135] M. Wang and R. M. Bradley, "Properties of GABAergic Neurons in the Rostral Solitary Tract Nucleus in Mice," *J. Neurophysiol.*, vol. 103, no. 6, Art. no. 6, Jun. 2010, doi: 10.1152/jn.00971.2009.
- [136] T. Okada, Y. Tashiro, F. Kato, Y. Yanagawa, K. Obata, and Y. Kawai, "Quantitative and immunohistochemical analysis of neuronal types in the mouse caudal nucleus tractus solitarius: Focus on GABAergic neurons," *J. Chem. Neuroanat.*, vol. 35, no. 3, Art. no. 3, May 2008, doi: 10.1016/j.jchemneu.2008.02.001.
- [137] F. C. Leitner *et al.*, "Spatially segregated feedforward and feedback neurons support differential odor processing in the lateral entorhinal cortex," *Nat. Neurosci.*, vol. 19, no. 7, Art. no. 7, Jul. 2016, doi: 10.1038/nn.4303.
- [138] A. Irintchev, M. Koch, L. K. Needham, P. Maness, and M. Schachner, "Impairment of sensorimotor gating in mice deficient in the cell adhesion molecule L1 or its close homologue, CHL1," *Brain Res.*, vol. 1029, no. 1, pp. 131–134, Dec. 2004, doi: 10.1016/j.brainres.2004.09.042.
- [139] P. E. Phelps, C. R. Houser, and J. E. Vaughn, "Immunocytochemical localization of choline acetyltransferase within the rat neostriatum: A correlated light and electron microscopic study of cholinergic neurons and synapses," *J. Comp. Neurol.*, vol. 238, no. 3, pp. 286–307, 1985, doi: 10.1002/cne.902380305.
- [140] V. V. Rymar, R. Sasseville, K. C. Luk, and A. F. Sadikot, "Neurogenesis and stereological morphometry of calretinin-immunoreactive GABAergic interneurons of the neostriatum," *J. Comp. Neurol.*, vol. 469, no. 3, pp. 325–339, 2004, doi: 10.1002/cne.11008.
- [141] K. C. Luk and A. F. Sadikot, "GABA promotes survival but not proliferation of parvalbumin-immunoreactive interneurons in rodent neostriatum: an in vivo study with stereology," *Neuroscience*, vol. 104, no. 1, pp. 93–103, Apr. 2001, doi: 10.1016/S0306-4522(01)00038-0.
- [142] O. Ibáñez-Sandoval, F. Tecuapetla, B. Unal, F. Shah, T. Koós, and J. M. Tepper, "A Novel Functionally Distinct Subtype of Striatal Neuropeptide Y Interneuron," *J. Neurosci.*, vol. 31, no. 46, pp. 16757–16769, Nov. 2011, doi: 10.1523/JNEUROSCI.2628-11.2011.
- [143] B. Ünal, F. Shah, J. Kothari, and J. M. Tepper, "Anatomical and electrophysiological changes in striatal TH interneurons after loss of the nigrostriatal dopaminergic pathway," *Brain Struct. Funct.*, vol. 220, no. 1, pp. 331–349, Jan. 2015, doi: 10.1007/s00429-013-0658-8.

# Dimitri Rodarie

PhD in computational neurosciences

+33 636356254  
d.rodarie@gmail.com  
57 rue Lafayette  
01200 Valserhône FRANCE  
25/04/1993

I am currently doing a thesis in Neurosciences at the Ecole Polytechnique Federale de Lausanne (EPFL) for the Blue Brain Project. My thesis subject deals with the development of a pipeline to systematically generate, simulate and refine point-neuron whole mouse brain models from biological data.

## EDUCATION

### Master's degree in Artificial Intelligence

*Université Claude Bernard Lyon 1*

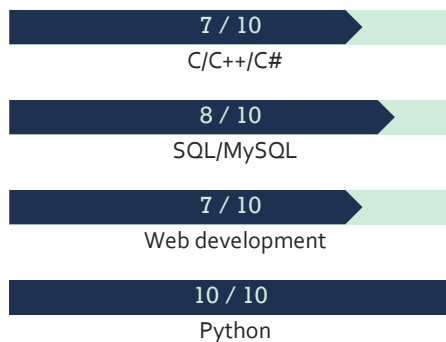
2015-2016

### Master's degree in computer sciences engineering

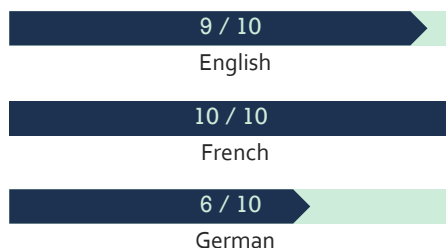
*Polytech Lyon Engineering School*

2011-2016

## SKILLS



## LANGUAGES



## INTERESTS

Music, Danse, Robotics

## EXPERIENCE

### Thesis in Computational Neurosciences

*Blue Brain Project / Geneva / 2017-2021*

Data-driven whole mouse brain modeling for closed-loop experiment

- Creation of a systematic workflow to generate, simulate, analyze, and refine whole mouse brain models at the level of point neurons.
- Development of a new approach to estimate cell type densities for the whole mouse brain.

### Master Project

*Human Brain Project / Geneva / 2016*

Research on models to control a simulated rodent musculoskeletal system in closed-loop to perform multi joint movements (Blender, Python).

- Modeling of a mouse musculoskeletal model in Blender.
- Optimization of movements using deep learning, genetic algorithms.

### Research internship

*LIRIS / Lyon / 2015*

Research on resources negotiation

- Development of strategy algorithms for multi-agent systems
- Creation of a negotiation framework